

Video Modeling and Noise Reduction in the Wavelet Domain

Nikhil Gupta

A Thesis

In the Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Electrical and Computer Engineering) at
Concordia University
Montreal, Quebec, Canada

December 2011

©Nikhil Gupta, 2011

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Nikhil Gupta

Entitled: Video Modeling and Noise Reduction in the Wavelet Domain

and submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (Electrical & Computer Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. G. Gouw

_____ External Examiner
Dr. W.B. Mikhael

_____ External to Program
Dr. T. Fancott

_____ Examiner
Dr. M.O. Ahmad

_____ Examiner
Dr. W.P. Zhu

_____ Co-Supervisor
Dr. E.I. Plotkin

_____ Co-Supervisor
Dr. M.N.S. Swamy

Approved by _____
Dr. J.X. Zhang, Graduate Program Director

November 28, 2011

Dr. Robin A.L. Drew, Dean
Faculty of Engineering & Computer Science

ABSTRACT

Video Modeling and Noise Reduction in the Wavelet Domain

Nikhil Gupta, Ph.D.

Concordia University, 2011

Digital video has become vastly popular in the last decade and is fast replacing its analog counterpart in every walk of life. However, like analog video, digital video is also plagued with additive noise which degrades its quality and hinders its processing. Hence, it is imperative that the corrupting noise be removed from the digital video. Although there are several methods to achieve this, not many techniques preserve the quality and the detail content of the video while filtering the noise aggressively.

This thesis concentrates primarily on the problem of additive noise removal from video sequences in the wavelet domain. A new statistical model for the subband coefficients of the video sequence is proposed. The spatial as well as the temporal subband data has been modeled using a prior based on the generalized Gaussian distribution. Along with modeling the a priori distribution of the spatial subband coefficients, this model accounts for the motion which occurs between successive frames.

A novel spatio-temporal filter for video denoising that operates entirely in the wavelet domain is then proposed. For effective noise reduction, the spatial and the temporal redundancies that exist in the wavelet domain representation of a video signal are exploited. The use of discrete cosine transform (DCT) is proposed to reduce the redundancies in the temporal direction. After the application of the DCT, the coefficients in the different wavelet domain subbands for the original

image sequence are modeled using a prior having a generalized Gaussian distribution. Based on this prior, filtering of the noisy wavelet coefficients in each subband is now carried out using a low-complexity wavelet shrinkage method that utilizes the correlation that exists between subsequent resolution levels.

Based on the proposed model, where the subband coefficients in individual frames as well as the wavelet coefficient difference occurring between two consecutive frames is modeled using the generalized Gaussian distribution, minimum mean squared error and maximum *a posteriori* Bayesian processors are proposed which estimate the noise-free wavelet coefficients in the current frame conditioned on the noisy coefficients in the current frame and the filtered coefficients in the past frame.

Based on the proposed statistical model, another novel noise reduction technique is proposed which exploits the spatial and the temporal redundancies that persist in the wavelet domain representation of the video sequence sequentially. The sequentially processed outputs of a Kalman filter and a spatial Bayesian filter are combined using an adaptive weighted averaging scheme.

The interscale dependencies in the subband representation of each frame are also modeled using a non-Gaussian bivariate distribution. The parameters for this bivariate distribution are estimated adaptively using the local correlations that exist between neighboring coefficients within each subband. Based on this bivariate distribution a shrinkage function is developed using the maximum *a posteriori* rule. To improve the performance of the filter, information from the adjacent frames is also incorporated in the shrinkage function.

Experimental results for all the presented algorithms show that the proposed schemes outperform several state-of-the-art spatio-temporal filters in terms of peak signal to noise ratio as well as visual quality.

ACKNOWLEDGEMENTS

It has been a great privilege for me to work with Dr. M. N. S. Swamy and Dr. Eugene Plotkin, my research advisers, who introduced me to signal processing in wavelet domain and then allowed me freedom to pursue research in video processing, a subject close to my heart. Exceptional teachers and researchers, they have been extra-ordinarily patient and supportive, having been always available for discussion and speedily responding to queries and research reports. I would like to take this opportunity to thank them for their continued encouragement and guidance through the course of my research.

I sincerely thank Dr. M. O. Ahmad and Dr. Weiping Zhu for insightful and timely discussions on research related problems at one point or another during the course of this thesis.

I extend my whole-hearted gratitude to my loving wife, Richa, and my parents, Vinod and Asha Gupta, and brother, Nishit Gupta, for their patience and constant encouragement and support, without which it would have been impossible to finish this work. Finally, I would like to thank my colleagues at CENSIPCOM and my friends for their valuable discussions, support and advice.

Nikhil Gupta, December 2011

I dedicate this work to my loving wife and family ...

Contents

List of Figures	x
List of Tables	xvi
List of Symbols	xix
List of Abbreviations	xxi
1 Introduction	1
1.1 Introduction and Motivation	1
1.2 Scope of the Thesis	4
1.3 Organization of the Thesis	5
1.4 Novelties and Contributions of This Thesis	8
2 Background and Literature Survey	9
2.1 Introduction to Wavelets	9
2.1.1 From Fourier Transform to Wavelets	10
2.1.2 Why Wavelets?	11
2.1.3 Wavelets and Multi-resolution Analysis	12
2.1.3.1 Multi-Resolution Analysis	12
2.1.3.2 Wavelet Functions	13

2.1.3.3	The Fast Wavelet Transform (FWT)	14
2.1.3.4	Orthogonal Wavelets	15
2.1.3.5	Bi-orthogonal Wavelets	17
2.1.4	Higher Dimensions and Translation Invariance	18
2.2	Literature Survey	21
2.3	Conclusion	29
3	Video Modeling	30
3.1	Introduction	30
3.2	Modeling of Subband Coefficients	32
3.3	Spatio-Temporal Modeling of Subband Coefficients	40
3.4	Conclusion	49
4	Wavelet Domain Video Denoising Using Temporal DCT and Adaptive Thresholding	51
4.1	Introduction	51
4.2	Problem Formulation	52
4.3	System Description	54
4.4	Temporal DCT	57
4.5	Hierarchically Adapted Wavelet Thresholding	58
4.6	Experimental Results and Discussion	69
4.6.1	Performance Criteria	69
4.6.2	Experimental Results	70
4.7	Conclusion	82
5	Bayesian Algorithms for Video Noise Reduction in the Wavelet Domain	84
5.1	Introduction	84

5.2	Problem Formulation	85
5.3	Spatio-temporal MMSE Estimation	87
5.4	Spatio-temporal MAP Estimation	92
5.5	Experimental Results and Discussion	97
5.6	Conclusion	105
6	Video Noise Reduction Using Adaptive Combination of Temporal and Spatial Filters	106
6.1	Introduction	106
6.2	Problem Formulation	107
6.3	Wavelet Domain Temporal Kalman Filter	109
6.4	Sequential Combination of Spatial and Temporal Estimates	111
6.5	Experimental Results and Discussion	113
6.6	Conclusion	119
7	Motion Adaptive Video Denoising Based on Bivariate Shrinkage of Wavelet Coefficients	122
7.1	Introduction	122
7.2	Problem Formulation and Bivariate Modeling	123
7.3	Adaptive Bivariate Shrinkage	126
7.4	Experimental Results and Discussion	130
7.5	Conclusion	136
8	Conclusion	138
8.1	Concluding Remarks	138
8.2	Scope for Further Work	141
	Bibliography	143

List of Figures

2.1	The filter bank algorithm for orthogonal wavelets: filtering and down-sampling of the signal I yields the low-pass signal LP and the high-pass signal HP. Signal S is reconstructed by up-sampling and filtering with the corresponding filters.	15
2.2	Orthogonal scaling and wavelet functions. (a) Daubechies scaling function with 4 vanishing moments, (b) Daubechies wavelet function with 4 vanishing moments, (c) Scaling function for the least asymmetric compactly supported wavelet, Symmlet, with 8 vanishing moments (Symmlet-8), and (d) Symmlet-8 wavelet function.	16
2.3	The filter bank algorithm for bi-orthogonal wavelets: filtering and down-sampling of the signal I yields the low-pass signal LP and the high-pass signal HP. Signal S is reconstructed by up-sampling and filtering with the corresponding filters.	18
2.4	Scaling and wavelet functions for the Bi-orthogonal 1-3 case. (a) Dual scaling function, (b) Dual wavelet function, (c) Scaling function, and (d) Wavelet function.	19
2.5	The two-dimensional wavelet transform (square variant) with two successive levels of decomposition.	19

2.6	The filter bank algorithm for a two-dimensional orthogonal wavelet transform of the two-dimensional signal I (e.g., image).	20
3.1	(a) Two consecutive frames of <i>Trevor</i> video sequence in time domain; (b) 3-level wavelet decomposition of the two frames in (a); and (c) The histogram of the coefficients corresponding to the HL_1 subband in the wavelet decomposition of two consecutive frames which are shown in (b).	33
3.2	Modeling of the subband coefficients generated using a 3-level decomposition of frame 20 of the <i>Trevor</i> video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.	38
3.3	Modeling of the subband coefficients generated using a 3-level decomposition of frame 10 of the <i>Train</i> video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.	39
3.4	(a) Two consecutive frames of <i>Trevor</i> video sequence in time domain; (b) HL_1 subband in the wavelet decomposition of two consecutive frames which are shown in (a). The subband coefficient values have been rescaled for visualization; and (c) The difference between the two subbands and the corresponding histogram. . . .	43
3.5	Modeling of the motion noise subband coefficients for frames 20 and 21 of the <i>Trevor</i> video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.	47

3.6	Modeling of the motion noise subband coefficients for frames 10 and 11 of the <i>Train</i> video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.	48
4.1	Block Diagram of the proposed video noise reduction scheme. . .	56
4.2	Subband structure and parent-child relationship in a 4-level orthogonal discrete wavelet transform	59
4.3	(a) Horizontal subbands from the wavelet and DCT transformed consecutive frames from the <i>Trevor</i> sequence; and (b) The corresponding histograms of the decorrelated subbands in (a)	61
4.4	Subbands HL_2 and HL_1 of a frame from sequence Miss America after application of DCT.	62
4.5	Modeling of the decorrelated subband coefficients for from the <i>Trevor</i> video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.	65
4.6	Modeling of the decorrelated subband coefficients from the <i>Train</i> video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.	66
4.7	PSNR comparison of the different algorithms for 60 frames of MA, Ha and PC sequences having input PSNRs of 20 dB, 25 dB and 25 dB, resp., for non-motion compensated filtering.	75
4.8	PSNR comparison of the different algorithms for 60 frames of MA, Ha and PC sequences having input PSNRs of 20 dB, 25 dB and 25 dB, resp., for Motion compensated filtering	77

4.9	PSNR comparison for MA and Ha sequences having input PSNRs of 20 dB and 25 dB, resp., using MC-DCWT for 2, 3 and 5 input frames	78
4.10	Filtered outputs of the 17 th frame of the <i>Miss America</i> sequence for non-motion-compensated filters	80
4.11	Filtered outputs of the 17 th frame of the <i>Miss America</i> sequence for motion-compensated filters	81
4.12	Filtered outputs of the 5 th field of the Hall sequence for proposed MC-DCWT filter with 2, 3 and 5 frames, respectively.	83
5.1	Shrinkage function obtained using the MMSE estimation of the current noise-free coefficients conditioned upon the current noisy coefficients and the denoised past frame coefficients. Also shown are the shrinkage operation as functions of the noisy subband coefficients in the current frame for a particular value of \hat{x}^{n-1}	90
5.2	Shrinkage function obtained using the MMSE estimation of the current noise-free coefficients conditioned upon the current noisy coefficients and the denoised past frame coefficients. Also shown are the shrinkage operation as functions of the noisy subband coefficients in the current frame for a particular value of \hat{x}^{n-1}	91
5.3	Shrinkage function obtained using the MAP estimation of the current noise-free coefficients conditioned upon the current noisy coefficients and the denoised past frame coefficients. Also shown are the shrinkage operation as functions of the noisy subband coefficients in the current frame for a particular value of \hat{x}^{n-1}	95

5.4	Shrinkage function obtained using the MAP estimation of the current noise-free coefficients conditioned upon the current noisy coefficients and the denoised past frame coefficients. Also shown are the shrinkage operation as functions of the noisy subband coefficients in the current frame for a particular value of \hat{x}^{n-1}	96
5.5	PSNR comparison of the different algorithms for <i>Miss America</i> sequence having an input PSNR of 20 dB	100
5.6	PSNR comparison of the different algorithms for <i>Hall</i> sequence having an input PSNR of 25 dB	101
5.7	PSNR comparison of the different algorithms for <i>Patrol Car</i> sequence having an input PSNR of 30 dB	102
5.8	Filtered (and zoomed) outputs of one frame of the <i>Patrol Car</i> sequence for the different filters	104
6.1	PSNR comparison of the different algorithms for <i>Hall</i> sequence having an input PSNR of 20 dB	116
6.2	PSNR comparison of the different algorithms for <i>Patrol Car</i> sequence having an input PSNR of 25 dB	117
6.3	PSNR comparison of the different algorithms for <i>Miss America</i> sequence having an input PSNR of 30 dB	118
6.4	Filtered outputs of the 17 th frame of the <i>Miss America</i> sequence for the different algorithms	120
7.1	(a) Empirical histogram obtained using the subband coefficients; (b) Histogram obtained using (7.4)	125
7.2	Bivariate shrinkage function obtained using the MAP estimation .	128

7.3	PSNR comparison of the different algorithms for <i>Patrol Car</i> sequence having an input PSNR of 20 dB	133
7.4	PSNR comparison of the different algorithms for <i>Miss America</i> sequence having an input PSNR of 25 dB	134
7.5	PSNR comparison of the different algorithms for <i>Hall</i> sequence having an input PSNR of 30 dB	135
7.6	Filtered (and zoomed) outputs of one frame of the <i>Patrol Car</i> sequence for the different filters	137

List of Tables

3.1	Average value of the generalized Gaussian distribution parameters, s_x and ν_x for the modeling of the DWT coefficients for some test videos for the LH subbands	36
3.2	Average value of the generalized Gaussian distribution parameters, s_x and ν_x for the modeling of the DWT coefficients for some test videos for the HL subbands	36
3.3	Average value of the generalized Gaussian distribution parameters, s_x and ν_x for the modeling of the DWT coefficients for some test videos for the HH subbands	37
3.4	Values of the metrics KSD and KLD for the modeling of the sub-band coefficients using GG, Gaussian and Laplacian distributions. The average values of KLD and KSD are obtained using all the frames of the 6 different test sequences.	41
3.5	Average value of the generalized Gaussian distribution parameters, s_m and ν_m for the modeling of the motion noise wavelet coefficients for some test videos for the LH subbands	45
3.6	Average value of the generalized Gaussian distribution parameters, s_m and ν_m for the modeling of the motion noise wavelet coefficients for some test videos for the HL subbands	45

3.7	Average value of the generalized Gaussian distribution parameters, s_m and ν_m for the modeling of the motion noise wavelet coefficients for some test videos for the HH subbands	45
3.8	Values of the metrics KSD and KLD for the modeling of the motion noise subband coefficients using GG, Gaussian and Laplacian distributions. The average values of KLD and KSD are obtained using all the frames of 6 different test sequences.	49
4.1	Average value of the generalized Gaussian distribution parameters, s_d and ν_d for the modeling of the decorrelated wavelet coefficients for some test videos for the LH subbands	63
4.2	Average value of the generalized Gaussian distribution parameters, s_d and ν_d for the modeling of the decorrelated wavelet coefficients for some test videos for the HL subbands	63
4.3	Average value of the generalized Gaussian distribution parameters, s_d and ν_d for the modeling of the decorrelated wavelet coefficients for some test videos for the HH subbands	63
4.4	Values of the metrics KSD and KLD for the modeling of the decorrelated subband coefficients using GG, Gaussian and Laplacian distributions. The average values of KLD and KSD are obtained using all the frames of 6 different test sequences.	67
4.5	PSNR comparison of non-motion-compensated and motion-compensated algorithms for the corrupted input test sequences with input PSNRs of 20 dB, 25 dB and 30 dB, resp.	72
4.6	PSNR comparison of the proposed non-motion-compensated and motion-compensated algorithms with 2, 3 and 5 corrupted input frames with input PSNRs of 20dB, 25 dB and 30 dB, resp.	76

5.1	PSNR comparison of the proposed algorithms, TMMSE and TMAP, with several others for the corrupted input test sequences with input PSNRs of 20 dB, 25 dB and 30 dB, resp.	99
6.1	PSNR comparison of the proposed algorithm, JKMMSE, with several others for the corrupted input test sequences with input PSNRs of 20 dB, 25 dB and 30 dB, resp.	114
7.1	PSNR comparison of the proposed algorithm, TBMAP, with several others for the corrupted input test sequences with input PSNRs of 20 dB, 25 dB and 30 dB, resp.	132

List of Symbols

$\phi(x)$:	scaling or father function
$\psi(x)$:	wavelet or mother function
$\tilde{\phi}(x)$:	dual scaling or father function
$\tilde{\psi}(x)$:	dual wavelet or mother function
h_k	:	low-pass filter coefficients
g_k	:	high-pass filter coefficients
\tilde{h}_k	:	dual low-pass filter coefficients
\tilde{g}_k	:	dual high-pass filter coefficients
$\langle \cdot, \cdot \rangle$:	inner product
$p_y(y)$:	probability density function of the random variable y
\mathbf{f}^n	:	matrix of discrete pixel values for the n^{th} video frame in time domain
η	:	matrix of additive noise in time domain
ϵ	:	matrix of additive noise in wavelet domain
\mathbf{g}	:	matrix of corrupted pixel values in time domain
σ_η	:	standard deviation of the corrupting additive white Gaussian noise
$\sigma_x, \hat{\sigma}_x$:	standard deviation of the noise-free coefficients in a subband and its estimate
σ_Y	:	standard deviation of the noisy coefficients in a subband
κ_x	:	kurtosis of the random variable x

ν_x, s_x	:	model parameters for the GG distribution for random variable x
$\mathbf{W}, \mathbf{W}^{-1}$:	two dimensional discrete wavelet transform matrix and its inverse
$\mathbf{X}^n, \hat{\mathbf{X}}^n$:	matrix of wavelet coefficients of the original video frame and its estimate
\mathbf{M}^n	:	motion noise
\mathbf{D}^n	:	matrix of DCT transformed wavelet coefficients
\mathbf{Y}^n	:	matrix of wavelet coefficients of the corrupted video frame
\mathbf{P}	:	matrix of parent coefficients
$\tilde{\mathbf{X}}$:	soft-thresholded matrix of coefficients
HAT	:	threshold parameter value
T_{bayes}	:	threshold estimated as per the <i>BayesShrink</i> rule
K, L	:	number of rows and columns in an image frame
N_s	:	total number of coefficients in a subband
γ_Z^n	:	correlation of the random variables in set \mathbf{Z}^n
$\hat{\mathbf{X}}_s^n$:	spatial estimate of the subband for JKMMSE filter
$\hat{\mathbf{X}}_k^n$:	temporal estimate for the JKMMSE filter
w	:	spatio-temporal combination weight for JKMMSE filter

List of Abbreviations

ATM	:	α -trimmed mean
AR	:	auto-regressive
AWA	:	adapted weighted averaging
AWGN	:	additive white Gaussian noise
CCD	:	charge coupled device
CDF	:	Cohen , Daubechies and Feauveau
DCT	:	discrete cosine transform
DCWT	:	discrete cosine-wavelet transform (proposed filter)
DWT	:	discrete wavelet transform
FIR	:	finite impulse response
Fo	:	Foreman video sequence
FWT	:	fast wavelet transform
GGD	:	generalized Gaussian distribution
GSM	:	Gaussian scale mixture
Ha	:	Hall video sequence
HAT	:	hierarchically adapted threshold
HVS	:	human visual system
iid	:	independent and identically distributed
IDWT	:	inverse Discrete Wavelet Transform

JKMMSE	:	joint Kalman minimum mean squared error (proposed filter)
JKW	:	joint Kalman Wiener
KLD	:	Kullback-Leibler Divergence
KSD	:	Kolmogorov-Smirnov Distance
KNN	:	K - nearest neighbors
LMMSE	:	linear minimum mean squared error
MA	:	Miss America video sequence
MAP	:	maximum <i>a posteriori</i>
MC-DCWT	:	motion-compensated discrete cosine-wavelet transform (proposed filter)
MC3DWTr	:	motion compensated 3-dimensional wavelet transform
MCA	:	multiple class averaging
MLE	:	maximum likelihood estimate
MMSE	:	minimum mean squared error
MRA	:	multi-resolution analysis
MRF	:	Markov random field
MSE	:	mean squared error
PC	:	Patrol Car video sequence
pdf	:	probability density function
PSNR	:	peak signal to noise ratio
RF	:	Rational filter
SNR	:	signal to noise ratio
STFT	:	short-time Fourier transform
SURE	:	Stein's unbiased risk estimate
TMAP	:	Temporal maximum <i>a posteriori</i> (proposed filter)
TMMSE	:	Temporal minimum mean squared error (proposed filter)
Tr	:	Train video sequence

Chapter 1

Introduction

1.1 Introduction and Motivation

There has been a significant increase in the popularity of the digital representation of image sequences (video) owing to the offered high quality, and the ease of transmission, processing and archiving. The digital form of video is fast replacing the analog form in the consumer and commercial markets and is already the preferred choice for data representation for entertainment as well as medical and scientific applications.

Digital video can be derived from quantization of the analog video sequence or can be produced directly from the scene. In either case, there are several factors which contribute to the degradation of these video sequences. Various degradations in image sequences can be classified [1] as follows:

- **Spatial degradations:** These can be due to the lens errors, defocus, non-zero dimensions of picture elements used in the imaging device, or bandwidth limitations in electronics. The effects are usually referred to as blur, unsharpness, ringing, echo, etc.

- **Temporal degradations:** These are caused by non-zero exposure time of the photosensitive material of the imaging device, by long decay times of light emitting materials, e.g., phosphors in displays and include motion blur and temporal flickering artifacts.
- **Geometrical degradations:** These are apparent as the distorted geometry of the displayed pictures, and are caused due to lens aberrations, deflection non-linearity in camera and display tubes.
- **Point-wise degradations:** These modify the grey-level of image sequence pixels and are referred to as noise, which is often some kind of random process, e.g., thermal noise, quantum noise etc., depending on the source of degradation.

In this thesis we focus on point-wise degradations, i.e., we consider noise in image sequences. Video sequences can be distorted by such noise during acquisition, recording and/or transmission. While acquiring the digital video from analog medium, noise can be introduced by the analog medium itself, the transmission channel, or the digitization process. The film and video-tape material on which the analog video is recorded is degraded with repeated use over time and results in the introduction of noise and other artifacts, which, in turn, are reproduced in the digitized form of the analog video. The process of digitization itself also introduces artifacts in the form of quantization noise. When the digital video is produced directly from the scene, noise is introduced due to the limitations of the imaging sensors. The scene being captured also has an impact on the quantity of noise introduced into the captured video sequence. For instance, in applications such as video surveillance, the low lighting conditions and the low contrast in the captured scene allows for more noise to be introduced. Other fast growing

applications, such as vision-based driver assistance systems, require that the same low-cost sensor be used to capture the entire gamut of dynamic ranges that are visible at all times of day. While such video sequences have a low noise content for ideal capture situations, such as sunny daytime conditions, a large amount of noise is introduced during dusk and nighttime conditions.

The corrupting noise typically results in degradation of the visual quality of the images in the sequence and may also mask important image information. Even if the perceived video at full-speed does not show noise degradation due to the temporal masking effect of the human visual system, common video analysis tasks, such as segmentation and tracking, might suffer in the presence of noise. In applications such as intelligent headlight control, where the detection of headlights or taillights covering an area less than ten pixels is important, the high noise levels can easily mask this information, causing the system to generate false positives or false negatives. Similar problems are encountered in automatic segmentation tasks in medical applications. Further, owing to its usually random nature, noise tends to raise the entropy of digital video, and therefore, hinders effective compression. Thus, it becomes imperative that the level of noise present in digital videos be reduced prior to any further processing or compression.

In many video applications the noise can be well approximated by the additive white Gaussian model [1], which we consider in this thesis. Removing noise from data can be considered as the process of constructing optimal estimates of the unknown signal or image sequence from the available noisy data. Video sequences can be considered as spatio-temporal data and in general, their denoising imposes a compromise between noise reduction and preserving significant visual details. In order to efficiently denoise such data and also avoid blurring during denoising, we not only need to make use of the spatial correlation that exists between the

pixels, but also of their temporal correlations or the correlation that exists between different frames in time. Due to the inherent dynamic nature of the video sequences, these signals are non-stationary in the temporal direction and trying to exploit temporal correlation during video noise reduction not only makes the video denoising different from the still image denoising, but also much more complex. Furthermore, for the denoising algorithm to have any practical application and for it to be realizable on embedded platforms for real-time use, its computational complexity as well as memory requirement has to be kept to a minimum.

In the last two decades, the wavelet transform [2]-[9] has emerged as a powerful tool that forms a bridge between theory and applications, and provides a suitable representation for processing data, that has enabled it to emerge as an important tool in denoising applications. The wavelet representation naturally allows for the construction of spatially adaptive algorithms that can preserve edges in a signal. It compresses the essential information in a signal into a few, large coefficients which represent the signal details at different resolution scales and facilitates the removal of the corrupting noise. In addition, while viewing images or video sequences, the human visual system (HVS) also employs multi-resolution decomposition similar to the wavelet decomposition to process the visual cues.

1.2 Scope of the Thesis

The discussion provided in the previous section serves to establish the fact that noise corrupts all kinds of videos: natural, scientific or medical. This corruption of the sequences hinders further processing (such as detection, segmentation, tracking, etc.) and compression. Thus, focus of this research is to develop new algorithms for video noise reduction that operate fully in the wavelet domain.

Unlike the common denoising algorithm development paradigm where a different framework is defined for each denoising algorithm, this thesis defines a unified development framework which not only models the subband spatio-temporal information corresponding to the video coefficients, but which could also be used for the development of several different noise reduction algorithms. Thus, the aim is to exploit the spatio-temporal redundancies in the video sequence not in an *ad-hoc* manner, but in a structured manner based on a statistical prior for the spatio-temporal data and develop noise reduction algorithms in the wavelet domain based on the single unified modeling framework developed in this research. Efforts have been made to keep the computational complexity of the denoising algorithms low so as to make their practical implementation feasible.

Since additive white Gaussian noise is the most common type of noise that corrupts the video sequences and is also studied extensively in literature, we primarily consider this kind of noise in this thesis.

We try to provide solutions to these problems that yield better results than the ones already existing in the literature while maintaining low computational complexity and real-time application. In doing so, we merge theory and practice, and employ heuristics too.

1.3 Organization of the Thesis

This thesis is organized as follows. In Chapter 2, a critical review of the existing literature on video noise reduction, both in the time and wavelet domains, has been provided. First, a detailed survey of the existing techniques for noise reduction is provided. Following this, the necessary background knowledge on wavelet theory is also reviewed. Attempt is made to present a concise, yet self-contained review

emphasizing the most relevant aspects for the topic of this thesis. We describe the benefits of the use of wavelets over the Fourier transform for the purpose of image denoising and compression. The multi-resolution analysis (MRA) of the input signal using wavelets is presented. We also describe the fast version of the discrete wavelet transform that is commonly employed in signal processing, using both orthogonal and bi-orthogonal wavelet families.

In Chapter 3, a statistical, spatio-temporal model is described for the video data in the wavelet domain, which allows for the motion to be captured and modeled using wavelet coefficients. This chapter describes the spatio-temporal relationship that exists between the wavelet coefficients belonging to the corresponding subbands in consecutive frames. The modeling of this relationship using a statistical distribution is also discussed. The simulation results of the modeling are presented and analyzed for their accuracy and efficacy.

Chapter 4 presents a novel spatio-temporal filter for video denoising that operates entirely in the wavelet domain and is based on exploiting the decorrelation of the spatial redundancies that exist in the wavelet domain representation of a video signal. We propose the use of discrete cosine transform (DCT) to reduce the redundancies in the temporal direction. After the application of the DCT, based on the *a priori* model discussed in Chapter 3, filtering of the noisy wavelet coefficients in each subband is described using a low-complexity wavelet shrinkage method that utilizes the correlation that exists between subsequent resolution levels. The experimental results validating the performance of the proposed algorithm are also presented.

In Chapter 5, we propose conditional Bayesian operators, which estimate the noise-free wavelet coefficients in the current frame conditioned on the noisy coefficients in the current frame and the filtered coefficients in the past frame. The

Bayesian operators using the minimum mean squared error (MMSE) criteria as well as the maximum *a posteriori* (MAP) criteria are utilized. Both the operators are adaptive to the spatial detail in the subbands as well as the temporal redundancies that exist in the wavelet domain. The formulation of the Bayesian operators is based on the statistical modeling of the wavelet coefficients described in Chapter 3. Experimental results are presented showing that the proposed scheme outperforms several state-of-the-art spatio-temporal filters in terms of the peak signal to noise ratio and the visual quality.

Chapter 6 describes the noise reduction in the video sequence via sequential exploitation of the spatial and the temporal redundancies that persist in the wavelet domain representation of the video sequence. The sequentially processed results are combined using an adaptive weighted averaging scheme. The algorithm utilizes the model proposed in Chapter 3 for the spatio-temporal processing and experimental results evaluating the performance of the algorithm are also presented.

In Chapter 7, each frame in the video sequence is transformed to the wavelet domain using the dual-tree complex wavelet transform. The interscale dependencies in subband representation of each frame are modeled using a non-Gaussian bivariate distribution. The parameters for this bivariate distribution are estimated adaptively using the local correlations that exist between neighboring coefficients within each subband. Based on this bivariate distribution, a shrinkage function is developed using the maximum *a posteriori* (MAP) rule. To improve the performance of the filter, information from the adjacent frames is also incorporated in the shrinkage function.

In Chapter 8, we conclude this thesis by summarizing the results obtained and suggesting directions for future research.

1.4 Novelties and Contributions of This Thesis

The main novelties and contributions of this thesis are: (i) a spatio-temporal video model which allows for *a priori* modeling of the video coefficients in the wavelet domain in the spatial as well as temporal orientation and (ii) five new algorithms for video noise reduction.

For a structured processing of the spatio-temporal information in the video sequences, a unified video model is proposed in this thesis. This model not only proposes the use of an *a priori* distribution for the modeling of the subband coefficients in the wavelet domain representation of each frame in the video sequence, but also allows for the modeling of the motion in terms of the corresponding wavelet coefficients in the consecutive subbands [14].

The proposed statistical video model can be used to denoise the videos using several different formulations. The five new algorithms proposed in this thesis for video noise reduction [10–16] are novel in the sense that they utilize the proposed statistical video model in low complexity frameworks and produce denoising results which are comparable or better than the current state-of-the-art while still being applicable for real-time application. The solutions provided exemplify the use of the video data in a multitude of ways, viz., decorrelating the temporal information before processing the spatial information, using the spatio-temporal information in a single Bayesian framework and processing the temporal as well as spatial information separately followed by adaptive combination of the results. These algorithms showcase the different ways the proposed model can be used for denoising video sequences using basic, low-complexity and adaptive techniques and can be extended to use more complex techniques in the same framework.

Chapter 2

Background and Literature Survey

This chapter first presents a brief overview of wavelets and multi-resolution analysis. Since the basis of this research is the processing of video sequences in the wavelet domain, an introduction to the preliminaries of wavelets is essential for completeness and understanding of this thesis. Followed by this overview, a comprehensive literature survey of the video noise reduction techniques is provided.

2.1 Introduction to Wavelets

This section provides a brief introduction to wavelets which can also be found in many books and papers at many different levels of exposition. Some of the standard books are [2–7]. Introductory papers include [8, 9, 17], and more technical ones are [18–20].

2.1.1 From Fourier Transform to Wavelets

The main idea behind wavelet analysis is to decompose a signal f into a basis of functions ψ_i :

$$f = \sum_i a_i \psi_i. \quad (2.1)$$

For an efficient decomposition of the signal f , a suitable family of the functions ψ_i is required that can represent the given signal using only a few coefficients a_i . These functions, ψ_i , should match the features of the signal to be decomposed.

Real-world signals usually are limited both in time (or space as in the case of images) and limited in frequency (band-limited). Time-limited signals can be represented efficiently using a basis of block functions (Dirac delta functions for infinitesimal small blocks). But block signals are not limited in frequency. Band-limited signals can be represented efficiently using a Fourier basis, but sines and cosines are not limited in time. The Fourier representation reveals the spectral content of a signal, but makes it impossible to recover the particular moment in time (or the particular space coordinates in case of images) where a certain change has occurred. This makes the Fourier representation inadequate when it comes to analyzing transient signals.

In signal, image and video processing, concentrating on transients (such as image discontinuities) is a strategy for selecting the most essential information from often an overwhelming amount of data. In order to facilitate the analysis of transient signals, i.e., to localize both the frequency and the time information in a signal, numerous transforms and bases have been proposed (see e.g., [4, 6]). Among these, in signal processing, the wavelet transform consists of functions that are both time-limited and band-limited. Wavelets are such functions which literally mean small waves, and have been termed so due to the requirements

that they should integrate to zero, waving above and below the axis, and that they should be localized. In wavelet analysis, the scale can be interpreted as the inverse of frequency. The wavelet transform acts as a microscope [6] focusing on smaller time phenomena as the scale decreases. This behavior permits a local characterization of signals, which the Fourier and the window Fourier transforms do not.

2.1.2 Why Wavelets?

Two of the main features of wavelets that are important for the applications that we shall consider are their good decorrelation and sparse representation of signals. Wavelets not only represent the signals sparsely by concentrating most of the energy in a few number of coefficients, they also act as edge detectors. They cluster the coefficients having significant amount of energy near one another. The locations of these clusters correspond to the locations of the spatial edges in the signal.

Some of the characteristics of wavelets that make them ideal for signal representation are as follows.

- Wavelets are localized in both the space/time and scale/frequency domains. Hence, they can easily detect local features in a signal.
- Wavelets are based on a multi-resolution analysis. Wavelet decomposition allows us to analyze a signal at different resolution levels (scales).
- Wavelets are smooth, which can be characterized by their number of vanishing moments. A function defined on the interval $[a, b]$ has n vanishing moments if

$$\int_a^b f(x)x^i dx = 0 \tag{2.2}$$

for $i = 0, 1, \dots, n - 1$. The number of vanishing moments represents how well smooth signals can be approximated in a wavelet basis.

Furthermore, there exist fast and stable algorithms to calculate the discrete wavelet transform and its inverse. The computational complexity of this fast algorithm is $O(n)$.

2.1.3 Wavelets and Multi-resolution Analysis

Multi-resolution analysis forms the basis for the development of discrete wavelets and hence, it is introduced first in this section.

2.1.3.1 Multi-Resolution Analysis

Consider the vector space \mathbf{L}^2 of the square integrable functions in \mathbf{R} :

$$\mathbf{L}^2 = \left\{ f : \int_{-\infty}^{+\infty} f^2(x) dx < \infty \right\}. \quad (2.3)$$

In a multi-resolution analysis, \mathbf{L}^2 is decomposed in nested subspaces V_j

$$\dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots \quad (2.4)$$

such that

1. $\bigcap_{j=-\infty}^{+\infty} V_j = \{0\}$ and $\overline{\bigcup_{j=-\infty}^{+\infty} V_j} = \mathbf{L}^2$
2. For any $f \in \mathbf{L}^2$ and any $j \in \mathbf{Z}$, $f(x) \in V_j$ if and only if $f(2x) \in V_{j-1}$
3. For any $f \in \mathbf{L}^2$ and any $k \in \mathbf{Z}$, $f(x) \in V_0$ if and only if $f(x - k) \in V_0$

If a function $\phi(x) \in V_0$, along with the set of its integer translates $\{\phi(x - k)\}_{k \in \mathbf{Z}}$, forms a basis for the space V_0 , it is called a *scaling function* or *father function*.

For the other subspaces V_j (with $j \neq 0$) we define:

$$\phi_{j,k}(x) \equiv 2^{\frac{j}{2}} \phi(2^j x - k). \quad (2.5)$$

2.1.3.2 Wavelet Functions

The nested nature of the subspaces allows the decomposition of V_{j+1} in terms of V_j and W_j , the orthogonal component of V_j in V_{j+1} :

$$V_j \oplus W_j = V_{j+1}, \quad (2.6)$$

$$W_j \perp V_j. \quad (2.7)$$

The direct sum of the subspaces W_j is equal to \mathbf{L}^2 :

$$\overline{\bigcup_{j=-\infty}^{+\infty} V_j} = \overline{\bigoplus_{j=-\infty}^{+\infty} W_j} = \mathbf{L}^2. \quad (2.8)$$

Thus, V_j is a “coarse-resolution” representation of V_{j+1} , while W_j carries the “high-resolution” difference between V_j and V_{j+1} .

If any function $\psi(x) \in W_0$ obeys the translation property (Property 3, Section 2.1.3.1) and along with its integer translates, $\{\psi(x - k)\}_{k \in \mathbf{Z}}$, forms a basis for the space W_0 , it is called a *wavelet function* or *mother function*. For the other subspaces W_j (with $j \neq 0$) we define:

$$\psi_{j,k}(x) \equiv 2^{\frac{j}{2}} \psi(2^j x - k). \quad (2.9)$$

2.1.3.3 The Fast Wavelet Transform (FWT)

Because V_0 and W_0 are both subspaces of V_1 , we can express $\phi(x)$ and $\psi(x)$ in terms of the basis functions of V_1 :

$$\phi(x) \equiv 2 \sum_k h_k \phi(2x - k), \quad (2.10)$$

$$\psi(x) \equiv 2 \sum_k g_k \phi(2x - k). \quad (2.11)$$

Due to multi-resolution analysis, these relations are valid between V_{j+1} , V_j and W_j for any arbitrary j as well. In the above relations, h_k and g_k are the low-pass and high-pass filter coefficients, respectively, that define the scaling function $\phi(x)$ and the wavelet function $\psi(x)$.

We can also express a function $f(x)$ that is written in terms of the basis functions of V_{j+1} in terms of the basis functions of V_j and W_j :

$$\begin{aligned} f(x) &= \sum_k \lambda_{j+1,k} \phi_{j+1,k}(x), \\ \Rightarrow f(x) &= \sum_l \lambda_{j,l} \phi_{j,l}(x) + \sum_l \gamma_{j,l} \psi_{j,l}(x). \end{aligned} \quad (2.12)$$

The transform coefficients, $\lambda_{j,l}$ and $\gamma_{j,l}$, are defined by:

$$\lambda_{j,l} = \sqrt{2} \sum_k h_{k-2l} \lambda_{j+1,k}, \quad (2.13)$$

$$\gamma_{j,l} = \sqrt{2} \sum_k g_{k-2l} \lambda_{j+1,k}, \quad (2.14)$$

This operation has a complexity $O(n)$ and is used to compute the discrete version of the wavelet transform (DWT) and hence, is known as the *Fast Wavelet Transform* (FWT). It is also known as the filter bank algorithm (Figure 2.1). The

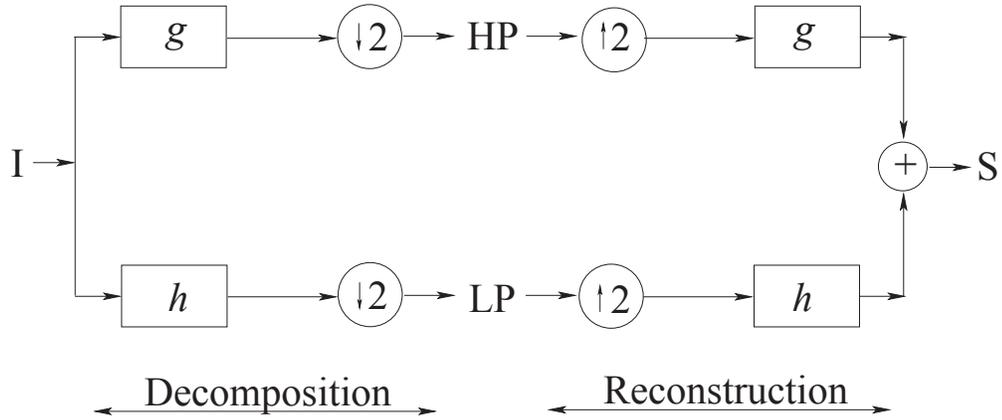


Figure 2.1: The filter bank algorithm for orthogonal wavelets: filtering and down-sampling of the signal I yields the low-pass signal LP and the high-pass signal HP. Signal S is reconstructed by up-sampling and filtering with the corresponding filters.

inverse wavelet transform can be obtained in a similar way.

2.1.3.4 Orthogonal Wavelets

If the $\phi_{j,k}$ and $\psi_{j,k}$ are orthonormal, i.e.,:

$$V_j \perp W_j, \quad (2.15)$$

$$\langle \phi_{j,l}, \phi_{j,l'} \rangle = \delta_{l-l'}, \quad (2.16)$$

$$\langle \psi_{j,l}, \psi_{j',l'} \rangle = \delta_{j-j'} \delta_{l-l'}, \quad (2.17)$$

then we can calculate the coefficients of the decomposition in (2.13) by taking the inner product of the function with scaling and wavelet functions:

$$\lambda_{j,l} = \langle f, \phi_{j,l} \rangle, \quad (2.18)$$

$$\gamma_{j,l} = \langle f, \psi_{j,l} \rangle. \quad (2.19)$$

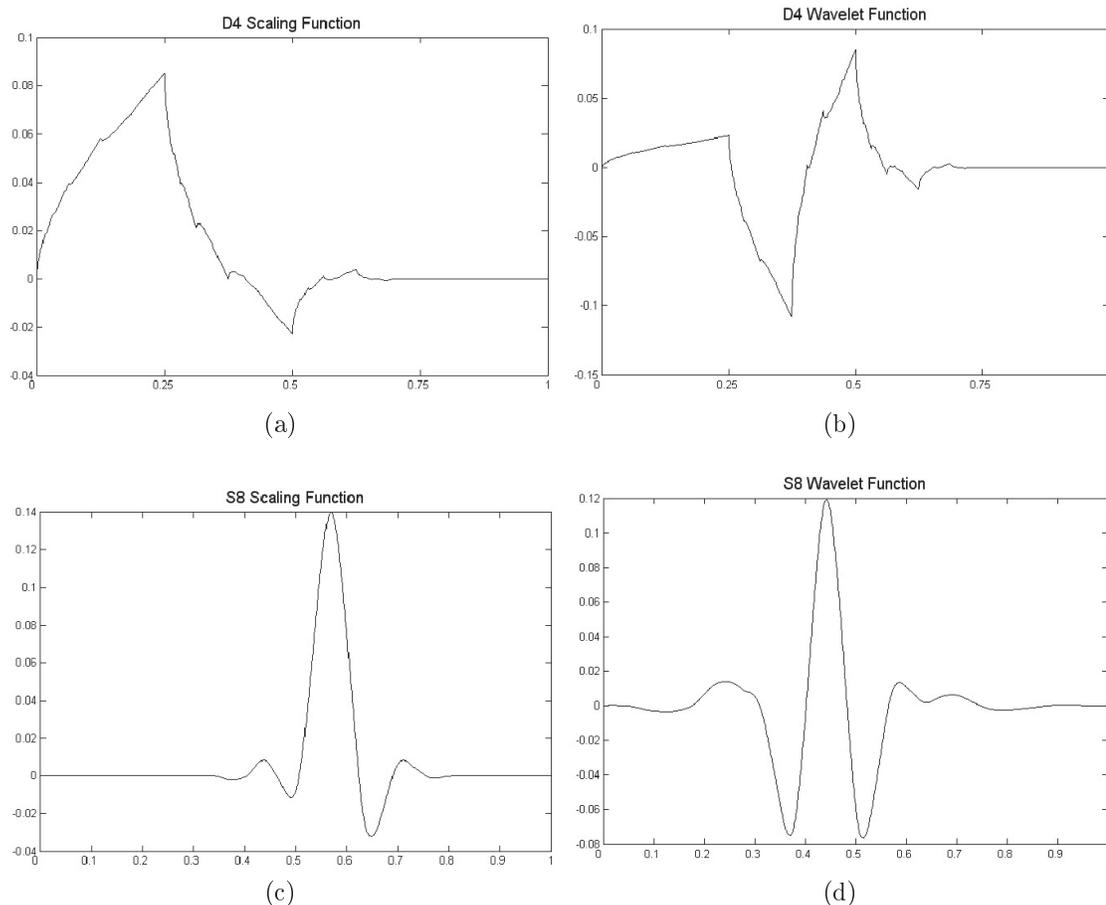


Figure 2.2: Orthogonal scaling and wavelet functions. (a) Daubechies scaling function with 4 vanishing moments, (b) Daubechies wavelet function with 4 vanishing moments, (c) Scaling function for the least asymmetric compactly supported wavelet, Symmlet, with 8 vanishing moments (Symmlet-8), and (d) Symmlet-8 wavelet function.

Examples of orthogonal wavelets are the family of orthogonal wavelets constructed by Daubechies [3]. The scaling function and the wavelet function for the member with 2 vanishing moments (also known as “D4” because the corresponding wavelet filter has 4 taps or filter coefficients) are shown in Figures 2.2(a) and (b), respectively. The Symmlet-8 is the least asymmetric compactly supported wavelet with 8 supporting moments and the corresponding scaling function and wavelet function are shown in Figures 2.2 (c) and (d), respectively.

2.1.3.5 Bi-orthogonal Wavelets

To obtain some special properties otherwise not available with the orthogonality conditions, such as linear phase and symmetry, we use the bi-orthogonality conditions in which we have two multi-resolution analyses, a primal and a dual:

- Primal: $V_j, W_j, \phi_{j,k}, \psi_{j,k}$.
- Dual : $\tilde{V}_j, \tilde{W}_j, \tilde{\phi}_{j,k}, \tilde{\psi}_{j,k}$.

The bi-orthogonality conditions imply:

$$V_j \perp \tilde{W}_j, \quad (2.20)$$

$$\tilde{V}_j \perp W_j, \quad (2.21)$$

$$\langle \tilde{\phi}_{j,l}, \phi_{j,l'} \rangle = \delta_{l-l'}, \quad (2.22)$$

$$\langle \tilde{\psi}_{j,l}, \psi_{j',l'} \rangle = \delta_{j-j'} \delta_{l-l'}. \quad (2.23)$$

The coefficients of the decomposition in the bi-orthogonal wavelet basis are obtained by taking the inner product of the function with the *dual* scaling and wavelet functions:

$$\lambda_{j,l} = \langle f, \tilde{\phi}_{j,l} \rangle, \quad (2.24)$$

$$\gamma_{j,l} = \langle f, \tilde{\psi}_{j,l} \rangle. \quad (2.25)$$

We can still use the filter bank algorithm if we use the dual filter pair (\tilde{h}, \tilde{g}) (related to the dual multi-resolution analysis) for the decomposition and the primal filter pair (h, g) (related to the primal multi-resolution analysis) for the reconstruction. Such a filter bank algorithm is depicted in Figure 2.3.

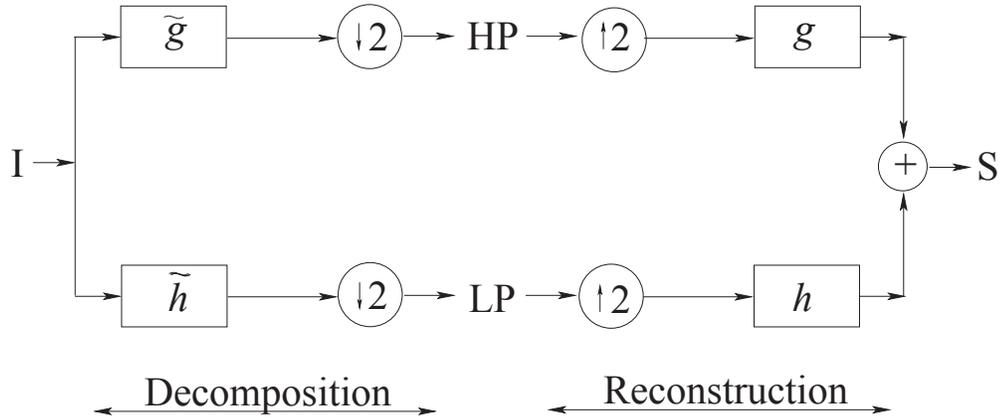


Figure 2.3: The filter bank algorithm for bi-orthogonal wavelets: filtering and down-sampling of the signal I yields the low-pass signal LP and the high-pass signal HP. Signal S is reconstructed by up-sampling and filtering with the corresponding filters.

Examples of bi-orthogonal wavelets are the family of bi-orthogonal wavelets constructed by Cohen, Daubechies and Feauveau (CDF) [21]. The primal and dual scaling and wavelet functions for the bi-orthogonal 1-3 member of the CDF family is illustrated in the Figure 2.4.

2.1.4 Higher Dimensions and Translation Invariance

The above wavelets were defined on a one-dimensional domain. To create wavelets for higher dimensional domains, we can perform the one-dimensional wavelet transform independently for each dimension in any order.

In the two-dimensional case we can get the square variant of the decomposition (see Figure 2.5) and the basis functions are the tensor products of the one-dimensional basis functions. After one transform step we have:

$$\frac{\phi(x) \otimes \phi(x) | \phi(x) \otimes \psi(x)}{\psi(x) \otimes \phi(x) | \psi(x) \otimes \psi(x)} \quad (2.26)$$

The filter bank algorithm for the two-dimensional wavelet transform for one

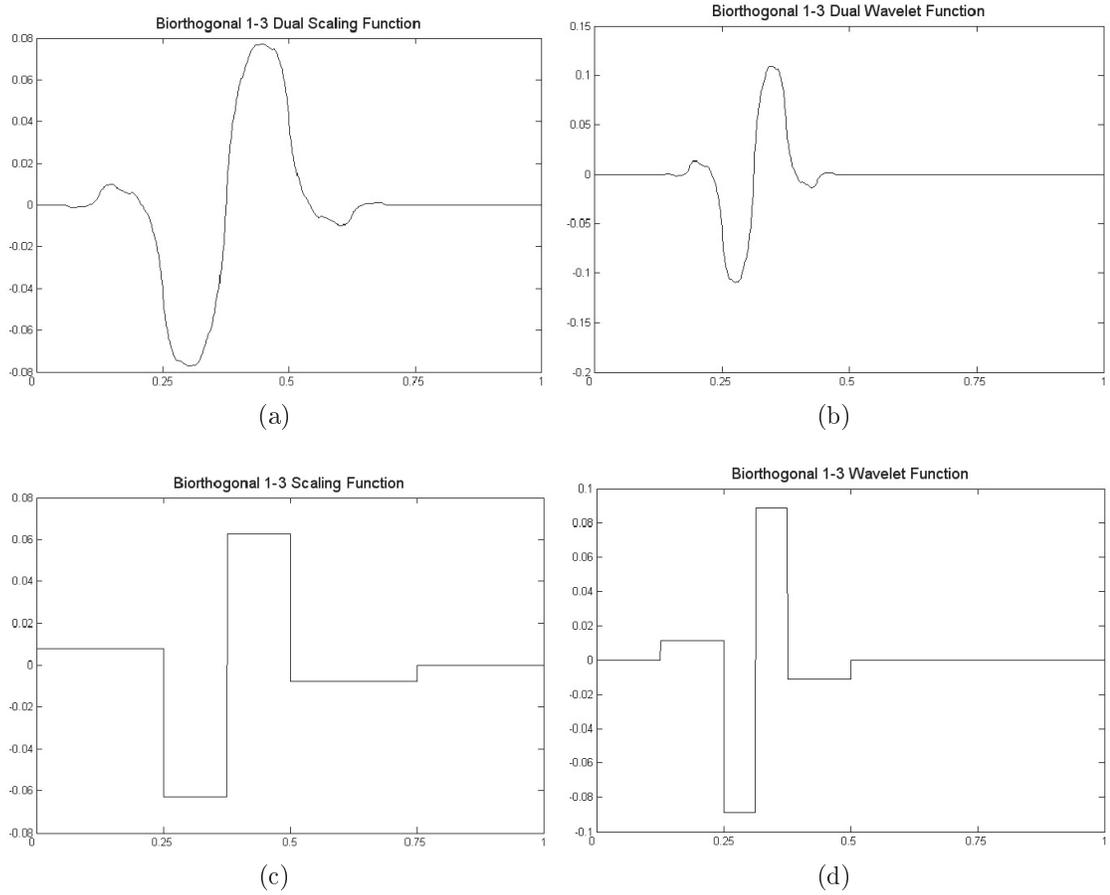


Figure 2.4: Scaling and wavelet functions for the Bi-orthogonal 1-3 case. (a) Dual scaling function, (b) Dual wavelet function, (c) Scaling function, and (d) Wavelet function.

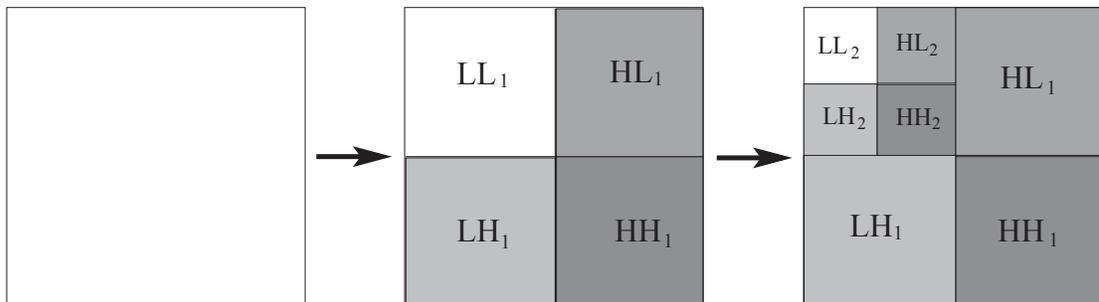


Figure 2.5: The two-dimensional wavelet transform (square variant) with two successive levels of decomposition.

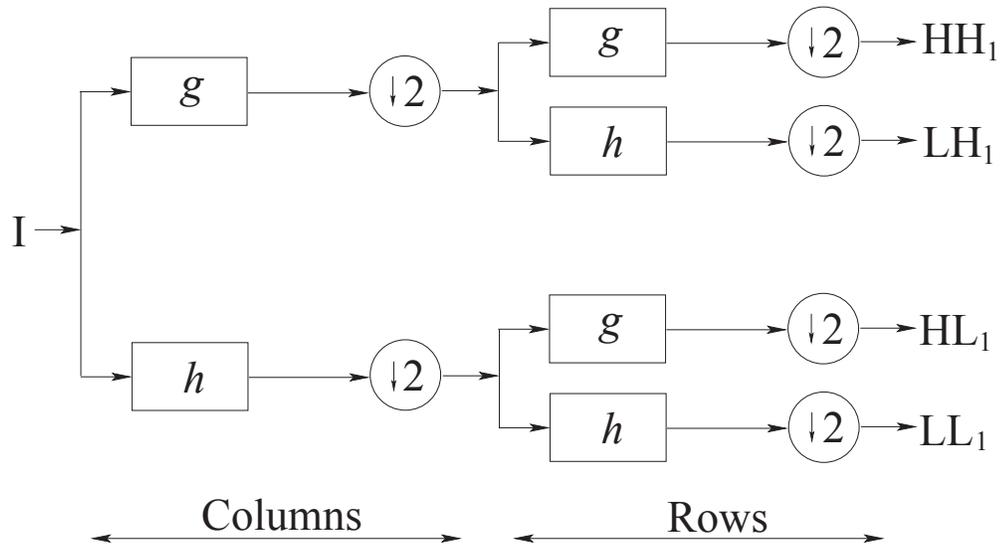


Figure 2.6: The filter bank algorithm for a two-dimensional orthogonal wavelet transform of the two-dimensional signal I (e.g., image).

decomposition level is shown in Figure 2.6. The image is decomposed into four subbands, namely LL , HL , LH and HH , where LL represents the *approximation subband* containing the low-pass equivalent of the image, HL represents the *detail subband* containing the horizontal edge information in the image, LH represents detail subband containing the vertical edge information in the image, and HH represents the detail subband containing the diagonal edge information in the image. At each subsequent level of decomposition, the approximation subband is further decomposed into four more subbands.

The sub-sampling of the signal in the filter bank causes the wavelet transform to vary with translations. Only if delay or advance is a multiple of 2^n , n being the number of transform levels, will the wavelet transform be a delayed or advanced version of the original transformed signal. In two dimensions, this condition needs to be true in both the vertical and horizontal directions.

The wavelet transform can be converted into an invariant transform by removing the sub-sampling step from the decomposition filter bank. This is called the

redundant wavelet transform or *stationary*, or *non-decimated wavelet transform* [22]. At each resolution step the filters have to be up-sampled to keep a consistent multi-resolution analysis. Also, the computational complexity is increased to $O(pn)$, with $p \leq \log(n)$ the number of transform levels. For the inverse transform, a mean value of appropriate transform coefficients has to be calculated due to the redundancy in the forward transform.

2.2 Literature Survey

In this section, a critical review of the existing literature pertaining to video noise reduction is provided. Compared to the extensive body of literature that exists for still image denoising, the work on noise reduction in video has been very limited. This can be attributed to a lack of good statistical models for video sequences and the immense computational power that is required to exploit the temporal correlation in videos. However, with digital videos gaining popularity and computational resources becoming cheaper, filtering of videos is fast becoming an important research area.

Image sequence data can be considered as spatio-temporal data. In order to efficiently denoise such data and also avoid blurring during denoising, we not only need to make use of the spatial correlation that exists between the pixels, but also of their temporal correlations, or the correlation that exists between different frames in time. Due to the inherent dynamic nature of the image sequences, these signals are non-uniform in the temporal direction and trying to exploit temporal correlation during video noise reduction not only makes the video denoising different from the image denoising, but also much more complex. Since temporal correlation is highest in the direction of the motion, most schemes in the litera-

ture employ either implicit or explicit motion estimation to effectively exploit the temporal correlation for noise reduction.

Based on the amount of motion utilized for the reduction of noise present in the video sequences, the existing filtering techniques can be divided into the following three classes:

- **Filters using no motion information.**

These filters employ no motion information and process the spatio-temporal video data as a set of three dimensionally arranged data to be filtered. In the class of such filters that use no motion information, 3D extensions of 2D non-linear filters, now employing a spatio-temporal support, have been proposed by Zlokolica *et al.* [23, 24]. The α -trimmed mean (ATM) filter [23] and the K - nearest neighbors (KNN) filter [24] are both based on order statistics of the pixels in the spatio-temporal window and in order to estimate the noise-free value of any pixel, they rely on the averaging of several neighboring pixels. In the case where motion is present in the video sequence, such filters cause severe blurring of edges and lead to insufficient filtering of noise.

- **Motion detected filters.**

In order to appropriately exploit the temporal correlation, most schemes in literature employ some form or the other of motion detection or estimation for noise reduction. The motion detected filters implicitly or explicitly detect motion in the video sequence. Such filters are designed based on a non-uniform model of the video sequence in the temporal direction and the use of the temporal data while filtering is dependent upon the extent of motion incurred in the sequence.

An example of the motion-detection-based approach is the filter proposed by Martinez and Lim [25], which uses a cascade of five 1D finite impulse response (FIR) linear minimum mean square error (LMMSE) estimators along five pre-hypothesized directions of motion. Such a scheme works well when the actual motion is along one of the predetermined directions, but leads to blurring in the presence of real motion which is usually more arbitrary. The directional filtering approach has been extended by Cocchia *et al.* in [26] by generalizing the 2D rational FIR filter proposed by Ramponi [27] to the 3D spatio-temporal region. The rational FIR filters have been applied along four spatial directions and nine temporal directions in conjunction with a movement detector to avoid “image flow” during fast motion. The spatial filtering in a spatio-temporal support has been further generalized by Jostschulte and Amer in [28] by detecting corners as well as edges (leading to eight spatial direction masks) and using an FIR filter to reduce noise in the spatial image region. The temporal filtering in [28] has been achieved by using a subband based scheme. The high pass channel employs a time recursive filter with frame delay, while the low pass channel employs a median based deinterlacing scheme that performs optimal filtering for different image content.

Zlokolica *et al.* have suggested a multiple class averaging (MCA) filter [29], where the pixels in the spatio-temporal support have been distributed into multiple classes and a weighted filter, dependent upon the class, prevalent motion and detail, has been used for averaging. Turney *et al.* [30] have modeled the image sequence as a temporal auto-regressive (AR) process and have suggested a low complexity, time recursive Kalman filter. Although this scheme, as all motion-detected schemes, provides good estimates in case of

stationary image sequences having low motion content, it either suspends temporal filtering or leads to blurring in the presence of motion.

- **Motion compensated filters.**

The blurring evident in the outputs generated by the filters that use no motion or only use motion detection can be minimized by estimating the motion and employing motion adaptive filtering. These kind of filters rely on explicit motion estimation and they remove the temporal non-uniformity of the image sequence using motion compensation. These filters are then designed for the uniform spatio-temporal data of the motion compensated image sequence .

In case of reliable motion estimates, simple temporal averaging [31] could also be used. However, motion vectors, more often than not, are not perfect. Thus, instead of simple averaging, Sezan *et al.* used an LMMSE filter in the temporal direction [32]. They have employed five pixels for the temporal window and this leads to large computational and memory requirements as the algorithm computes two motion vectors (forward and backward) for each pixel in addition to storing the two past and two future frames.

The computational and memory requirements are generally reduced by limiting the temporal support and employing a motion-compensated spatio-temporal neighborhood. The compensation of motion removes or reduces any non-stationarity that originally existed in the temporal direction. Furthermore, there exists a trade-off between the extent of the spatio-temporal support and the benefit gained in terms of noise reduction as having a bigger spatio-temporal neighborhood might, in turn, lead to blurring of the data at the center of the support window, which is being currently filtered.

Kokaram *et al.* [33] used a 3D Wiener filter on the motion-compensated and overlapped spatio-temporal support. Overlapping the blocks prevents the block artifacts across block edges. Samy [34] adapted the 3D Wiener filter to estimate the mean and variance locally over the spatio-temporal neighborhood. However, the filter in [34] does not account for the presence of outliers in the neighborhood. This approach was improved upon by Ozkan *et al.* in [35], where the LMMSE filter switches between the temporal and spatio-temporal neighborhood depending upon the uniformity of the pixels in the neighborhood. This idea was further generalized in the adapted weighted averaging (AWA) filter [35] by choosing the most uniform subset from the spatio-temporal neighborhood. Pixel difference has been used as the criteria for uniformity and the pixels included in the subset are weighted according to their difference from the center pixel in the averaging scheme. The AWA filter is particularly suited for low SNRs and abruptly changing scene content. Motion estimation in [35] used Fogel’s optical flow algorithm [36] and the hierarchical block matching algorithm proposed in [37].

Reduced-order, time recursive, Kalman filtering has been used in [38] to reduce the number of frames to be stored. The spatio-temporal neighborhood is considered as the state and the Kalman filter is applied along the motion trajectory. Although such a scheme is able to utilize the spatial as well as temporal correlation, the computational burden is still high. Dugad and Ahuja [39] have further reduced the computational burden by carrying out the spatial and temporal filtering sequentially. They have used an optimal Wiener filter for spatial filtering and a Kalman filter for temporal filtering (Joint Kalman Wiener (JKW) filter), and have reported good results for videos having varying amount of motion content.

All the motion-compensated noise reduction algorithms discussed above employ motion estimation as an integral part. This makes the performance of the motion-compensated filters dependent upon the algorithm used for motion estimation. This not only makes such algorithms computationally expensive, but also memory intensive. There is a need to store the two motion vectors for each past and future frame considered in the processing, along with the storage required for the frames themselves. Generally block motion estimation algorithms are used for motion-compensated video denoising as they provide a good compromise between accuracy and complexity. Various schemes for motion estimation and also for more efficient and effective search patterns for block matching algorithms have been presented in literature [36, 37, 40–44].

Along with the extent to which the different video noise reduction approaches utilize motion, they can also be categorized based on the domain they operate in. Almost all of the approaches discussed above process the video sequence in the original spatio-temporal domain. However, wavelets have emerged as a powerful multi-resolution tool for signal analysis in the last decade. In the case of still image denoising, wavelet thresholding has emerged as a preferred tool and been shown to outperform other means of denoising [45–51]. Their utility as a tool for video noise reduction has also been explored in recent years.

There are two broad approaches to exploit the wavelet representation for the spatio-temporal filtering of the noisy video data. Firstly, the coefficients of the 3-D wavelet representation of a noisy video could be thresholded to yield the noise-free estimates of the original video sequence [53–55]. There is no need for explicit motion detection, estimation or compensation in this approach. However, its main drawbacks are the large memory requirements which is constrained by the length

of the wavelet filter coefficients [56], the very large computational complexity necessary to process the 3-D block of data together and the long latency between the output of the two successive blocks of processed frames, making the filtering process discontinuous in time.

Dabov *et al.* [57] perform the local 3-D transform on a group of blocks found by using block-matching with respect to the one being currently processed. Following the 3-D transform, each 3D group in the subband domain is filtered and aggregated using a weighted average. Since it uses two separate steps involving the intermediate estimate production using hard thresholding followed by block-matching and Weiner filtering, this scheme requires a large amount of computational power. Furthermore, since the block-matching and denoising is performed on temporal groups of frames, the memory requirements and the latency in this technique are very high, thus rendering it unsuitable for any real-time applications. Yu *et al.* [58] have used the 3D DWT around any given frame by considering a group of neighboring frames and then first applying the 1D DWT transform in the temporal direction, followed by the 2D spatial DWT. The framework used in [58] is similar to that initially proposed in [10, 14] with one of the differences being that in [58] the past denoised coefficients are used as the neighbors to the current noisy frame as compared to the input noisy frame neighbors in [10, 14]. The performance gain in terms of the denoising improvement in [58] is obtained at the cost of additional memory access, which in turn has an impact on the real-time implementation of the technique.

The second approach is to consider the 2-D wavelet representation of each frame and to process the subsequent series of 2-D wavelet domain frames. Acknowledging the better framework provided by the wavelet domain for denoising or filtering, [29] has also extended the spatio-temporal filter to the wavelet domain.

However, this filter uses weighted averaging of neighboring coefficients to reduce the noise in the wavelet domain and produces blurring of the high frequency information and thus, leads to a loss of detail in the denoised image sequence. While the wavelet domain filter in [29] operates entirely in the wavelet domain, Pizurica *et al.* [52] have proposed a noise reduction scheme that utilizes wavelet domain for spatial filtering but reverts back to the time domain for exploiting the temporal correlation. However, the computational expense involved with [29, 52] makes them unsuitable for real- or near real-time implementation.

Balster *et al.* [59] have used a wavelet-based image denoising technique developed in [60] for spatial filtering of a noisy frame. Haar wavelet-based thresholding has been used on the spatially-denoised frames to perform the temporal noise reduction. In [56], the motion trajectory of the different coefficients has been estimated and a recursive temporal filtering is performed on the spatially filtered [61] coefficients along the estimated trajectories. Rahman *et al.* [62] proposed a joint probability density function to model the video wavelet coefficients of any two neighboring frames and then applied this statistical model for denoising. The joint density function has been used for spatial filtering of the noisy wavelet coefficients by developing a bivariate maximum *a posteriori* estimator.

As is evident from the above discussion, even though wavelets have been successfully employed for noise reduction in video sequences, most of the recent work has not utilized the spatio-temporal information to the fullest extent and a wide variety of *ad hoc* frameworks have been used for the development of these filters. The high latency, memory requirements and computational complexity have hindered the application of most of these techniques in real-time applications.

2.3 Conclusion

In this chapter, a brief overview of the wavelet transform has been provided as this forms the basis for the development of the spatio-temporal video modeling as well as the noise reduction filters in the following chapters. The reasons for the widespread use of wavelets as a key tool in signal processing have been explored and the properties of wavelets which make them useful, such as multi-resolution, orthogonality, linearity and sparsity have also been discussed.

Following the overview of the wavelet transform, a survey of the current noise reduction techniques that exist in the literature for video sequences has been provided.

Chapter 3

Video Modeling

3.1 Introduction

As noted in the previous chapter, wavelets have emerged as a powerful tool for the processing of images and videos for a variety of applications, such as denoising, compression, and watermarking. One of the main reasons for this popularity is the ability to use statistical approaches in the development of wavelet based techniques via the use of a prior function for the wavelet subband coefficients. While there has been a lot of work to describe the probability density function (p.d.f.) of the wavelet coefficients of 2D images, there has been little effort to develop specific models which capture the inherent nature of video sequences.

In the case of 2D images, the basic idea is to model the DWT coefficients of an image with a pre-specified probability density function. This modeling tries to exploit the nature of the histogram of the wavelet coefficients, which are characterized by a sharp central peak and a heavy tail [63, 64]. Some of the standard density functions for the 2-D DWT coefficients of images are the generalized Gaussian (GG) density function [65], the Bessel K-form density function [64], and the

alpha-stable density function [66]. While a common assumption in the use of any of these densities is the independent and identically distributed (i.i.d.) nature of the wavelet coefficients, the wavelet coefficients in different subbands are actually correlated to one another and this can also be exploited in the *a priori* description of the subband coefficients [67, 68]. This inter-scale dependency has been modeled via the use of spherically invariant bivariate p.d.f. in [67] and modeling the wavelet coefficients by the hidden Markov tree model [68].

For video processing, usually the same wavelet coefficient priors which were developed for 2D images are used to represent the sequences of 2D frames [56, 59, 61]. It is obvious that the way in which the statistical information represented by the subband coefficients is modeled using *a priori* models plays a significant role in the performance of the algorithms being developed. Thus, in this chapter we discuss the modeling of the spatio-temporal subband coefficients to capture the information inherent in the video sequence. To achieve this, the modeling of the wavelet coefficients corresponding the video sequence has been considered in two steps:

1. The modeling of the subband coefficients corresponding to the spatial 2D video frames. As discussed above, the modeling of the coefficients has an effect on the overall performance of the algorithm utilizing that p.d.f.. The more closely the *a priori* p.d.f. is able to model the sharp central peaks and heavy tails of the coefficient distribution, the better is the algorithmic performance. While the heavy tailed distributions [62, 64–66] are much better suited for this modeling as compared to several other commonly used distributions, such as Gaussian and Laplacian, only incremental improvements are achieved by using different heavy tailed distributions. Towards this end, the modeling of the subband coefficients corresponding to the 2D frames is

presented in this chapter.

2. The spatio-temporal modeling of the subband coefficients to account for the information inherent in the video sequences. Since the modeling of the subband coefficients corresponding to the 2D frames does not take into consideration the temporal information evident in the video sequences, this chapter also discusses a new model which takes into consideration the spatio-temporal information.

3.2 Modeling of Subband Coefficients

In this section, the modeling of the subband coefficients is discussed in detail.

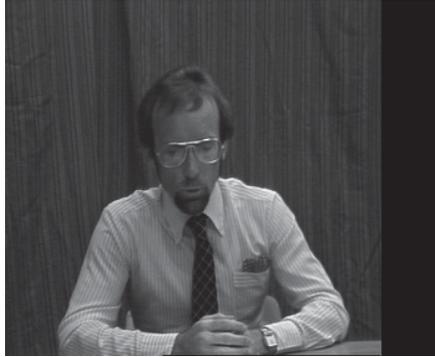
Consider a video sequence which has been discretized both in the spatial and temporal coordinates. All the pixels in the n^{th} frame of such a video sequence, example of which is shown in the Figure 3.1 (a), can be represented as \mathbf{f}^n :

$$\mathbf{f}^n = \{f_{ij}^n | i = 1, 2, \dots, L \text{ and } j = 1, 2, \dots, K\} \quad (3.1)$$

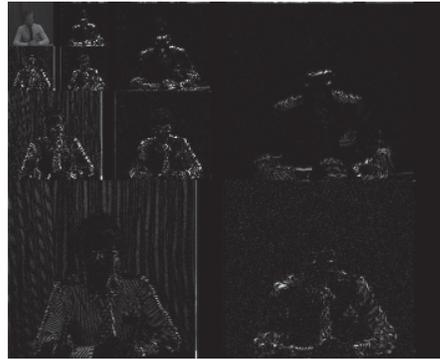
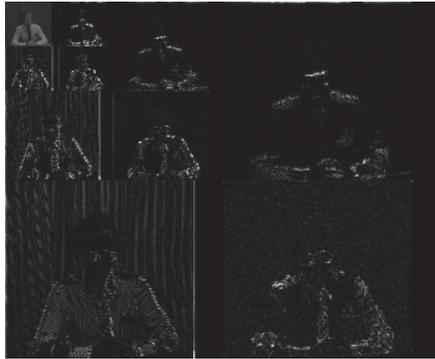
where (i, j) represent the row and column coordinates of the pixel, respectively, n denotes the frame index, and L and K are the number of rows and columns in each frame, respectively. The bold-faced notation is used to represent the set of pixels in the n^{th} frame.

Note that although 3.1 implies a 3-D sampling or a progressive scanning of the video sequence, the same model would be valid for an interlaced image sequence if the even and odd fields are considered as two separate frames in a progressively scanned image sequence.

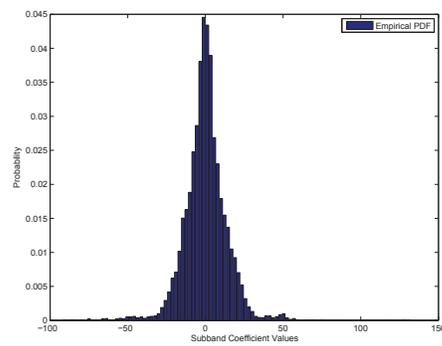
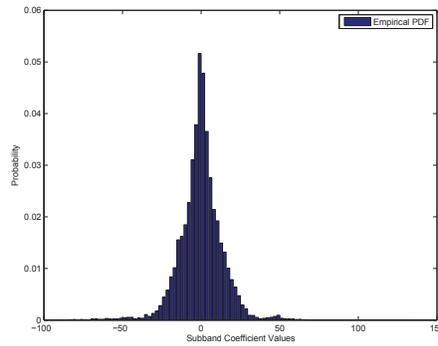
Each frame in a video sequence can be processed using the 2D discrete wavelet transform to yield the corresponding wavelet domain representation of the video



(a) Original Frames 20 and 21 from *Trevor* Sequence



(b) 3-level wavelet decomposition



(c) Histogram of the HL_1 subbands of two consecutive frames of *Trevor* sequence

Figure 3.1: (a) Two consecutive frames of *Trevor* video sequence in time domain; (b) 3-level wavelet decomposition of the two frames in (a); and (c) The histogram of the coefficients corresponding to the HL_1 subband in the wavelet decomposition of two consecutive frames which are shown in (b).

sequence. The corresponding coefficients in the wavelet domain can be obtained as

$$\mathbf{X}^n = DWT(\mathbf{f}^n) \quad (3.2)$$

where \mathbf{X}^n is the wavelet domain representation of the input frame n and can be represented as

$$\mathbf{X}^n = \{X_{ij}^n | i = 1, 2, \dots, L \text{ and } j = 1, 2, \dots, K\} \quad (3.3)$$

where, similar to the pixel domain representation, (i, j) represent the row and column coordinates of the wavelet coefficients, respectively, n denotes the transformed frame index, and L and K are the number of rows and columns in each wavelet domain representation, respectively. The bold-faced notation is used to represent the set of pixels in the n^{th} frame.

The wavelet decomposition of two successive frames from the *Trevor* video sequence is shown in Figure 3.1 (b). As discussed in Sections 2.1.3 and 2.1.4 and as is evident from this figure, the 2D wavelet decomposition of the input frame has a dyadic structure which is suited to multi-resolution analysis. Furthermore, owing to clustering, the wavelet significant coefficients tend to accumulate around the edges in the spatial domain. However, even this correlation is extremely short-lived and generally does not exist beyond the immediate neighbor. Because of this, the wavelet coefficients of images have been extensively modeled as independent and identically distributed random variables (e.g., [6, 45, 48, 49, 61, 65, 69–80]). Thus, in this thesis, the wavelet coefficients of the video sequence frames are modeled as i.i.d. variables.

Figure 3.1 (c) presents the histograms of the wavelet coefficients belonging to

the HL_1 subbands of the decomposition shown in Figure 3.1 (b). The histogram of the subband coefficients is symmetric, zero-mean, uni-modal, sharply peaked and heavy tailed in nature [63, 64]. Owing to these properties, the generalized Gaussian distribution has been used here to model the distribution of the frame subband coefficients.

The GG distribution prior that can be used to describe the pdf of the subband coefficients, which are treated as zero-mean, i.i.d. random variables, can be presented as

$$p_x(x) = \frac{\nu_x}{2s_x\Gamma(\frac{1}{\nu_x})} \exp(-|\frac{x}{s_x}|^{\nu_x}), \quad s_x, \nu_x > 0 \quad (3.4)$$

where $p_x(x)$ is the pdf of the random variable x , $\Gamma(a) = \int_0^\infty b^{a-1} \exp(-b) db$ is the Gamma function and the shape parameter, and ν_x and s_x are the model parameters. The variance of a GGD random variable can be defined as

$$\sigma_x^2 = s_x^2 \frac{\Gamma(\frac{3}{\nu_x})}{\Gamma(\frac{1}{\nu_x})}, \quad (3.5)$$

while the kurtosis can be expressed as

$$\kappa_x = \frac{\Gamma(\frac{1}{\nu_x})\Gamma(\frac{5}{\nu_x})}{\Gamma^2(\frac{3}{\nu_x})}. \quad (3.6)$$

For each subband, the model parameters ν_x and s_x can be estimated by using the variance and the kurtosis of the available subband coefficients; the shape parameter, ν_x , typically lies in the range (0, 2].

Two of the special cases of the above distribution which have been widely used in literature are:

- $\nu_x = 1$: In this case, the GG distribution reduces to Laplacian distribution with a zero mean and a variance of $2s_x^2$.

Test Image	ν_x			s_x		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
Miss America	0.6832	0.4915	0.6073	0.9609	0.6300	4.8961
Hall	0.4379	0.6042	0.7390	0.6118	7.2671	27.5986
Train	0.4415	0.4480	0.6372	0.6429	2.3000	14.8839
Patrol Car	0.3983	0.4276	0.6365	0.1729	0.6380	13.9702
Salesman	0.6799	0.8070	1.1213	2.3370	6.7293	15.8566
Trevor	0.7578	0.7363	0.8354	2.5948	5.9693	21.2024

Table 3.1: Average value of the generalized Gaussian distribution parameters, s_x and ν_x for the modeling of the DWT coefficients for some test videos for the LH subbands

Test Image	ν_x			s_x		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
Miss America	0.3125	0.4191	0.4699	0.0262	0.3954	2.0637
Hall	0.3895	0.5340	0.6103	0.3295	4.2860	16.5382
Train	0.6115	0.8368	0.9206	6.1105	22.0139	51.2772
Patrol Car	0.7076	0.7424	0.7986	7.4120	20.3449	54.5289
Salesman	1.1471	1.3493	1.5532	21.4858	53.0975	132.1720
Trevor	0.4044	0.4653	0.5533	0.1966	1.1095	5.7199

Table 3.2: Average value of the generalized Gaussian distribution parameters, s_x and ν_x for the modeling of the DWT coefficients for some test videos for the HL subbands

- $\nu_x = 2$: In this case, the GG distribution is equivalent to the Gaussian distribution which has a zero mean and a variance of $\frac{s_x^2}{2}$.

To evaluate the modeling of the empirical distribution by the GG distribution, six different video sequences have been used. All the frames in each of the video sequences have been transformed to the wavelet domain using a three level decomposition. The resulting wavelet coefficients corresponding to each of the subbands were then approximated using the GG distribution. The model parameters (ν_x and s_x) for each subband were estimated from its variance and kurtosis as described above. Tables 3.1, 3.2 and 3.3 present the estimated values of the

Test Image	ν_x			s_x		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
Miss America	1.0177	0.4255	0.5312	1.6216	0.1407	1.4561
Hall	0.3634	0.3968	0.5149	0.0380	0.3538	3.3652
Train	0.4959	0.6676	0.5063	0.3387	4.2267	3.7797
Patrol Car	0.3310	0.3459	0.3758	0.0151	0.0674	0.4877
Salesman	0.6860	0.9413	1.1324	2.4479	8.9292	16.1209
Trevor	0.5975	0.4694	0.5710	0.4004	0.4927	3.4140

Table 3.3: Average value of the generalized Gaussian distribution parameters, s_x and ν_x for the modeling of the DWT coefficients for some test videos for the HH subbands

model parameters for each of the subbands for the six videos used for evaluation. The results are obtained by averaging the model parameters for corresponding subbands over all the frames of the respective video sequences. To compare the relative performance of this approximation, the same empirical subband data was also modeled using the Gaussian and Laplacian distributions. Figures 3.2 and 3.3 show the fitted histograms of all the subbands belonging to frame 20 of *Trevor* and frame 10 of *Train* video sequences, respectively. The efficacy of the GGD prior to model the subband coefficients much better than the Gaussian and Laplacian distributions is evident in these histograms. The GGD prior is able to closely model the sharp peak in the histograms around the zero coefficient magnitude as well as the heavy tails.

To quantify the performance of the different PDF priors in fitting the empirical PDF the following two metrics have been used:

- Kolmogorov-Smirnov Distance (KSD): The KSD quantifies a distance between the empirical distribution function and the reference distribution used

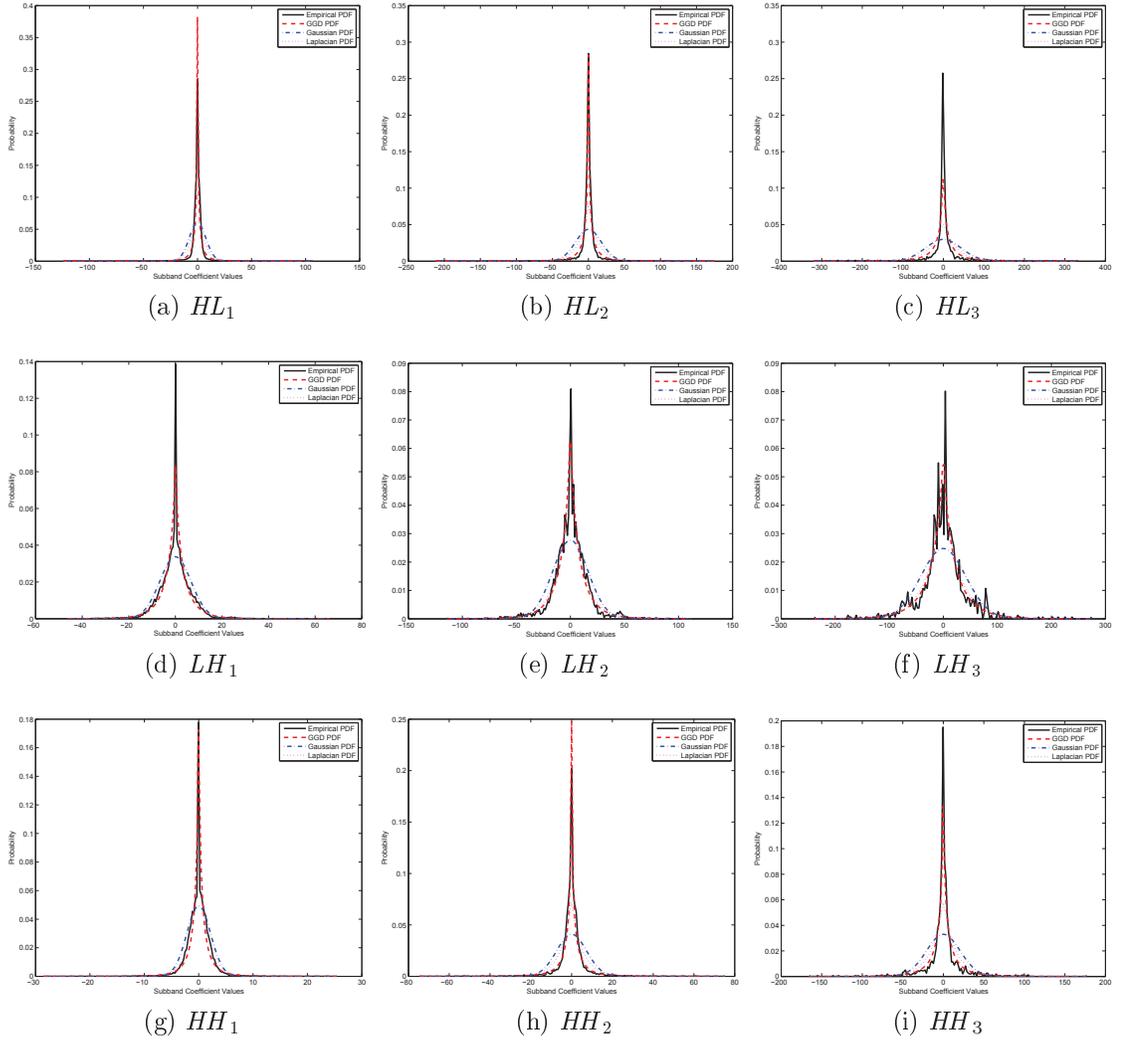


Figure 3.2: Modeling of the subband coefficients generated using a 3-level decomposition of frame 20 of the *Trevor* video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.

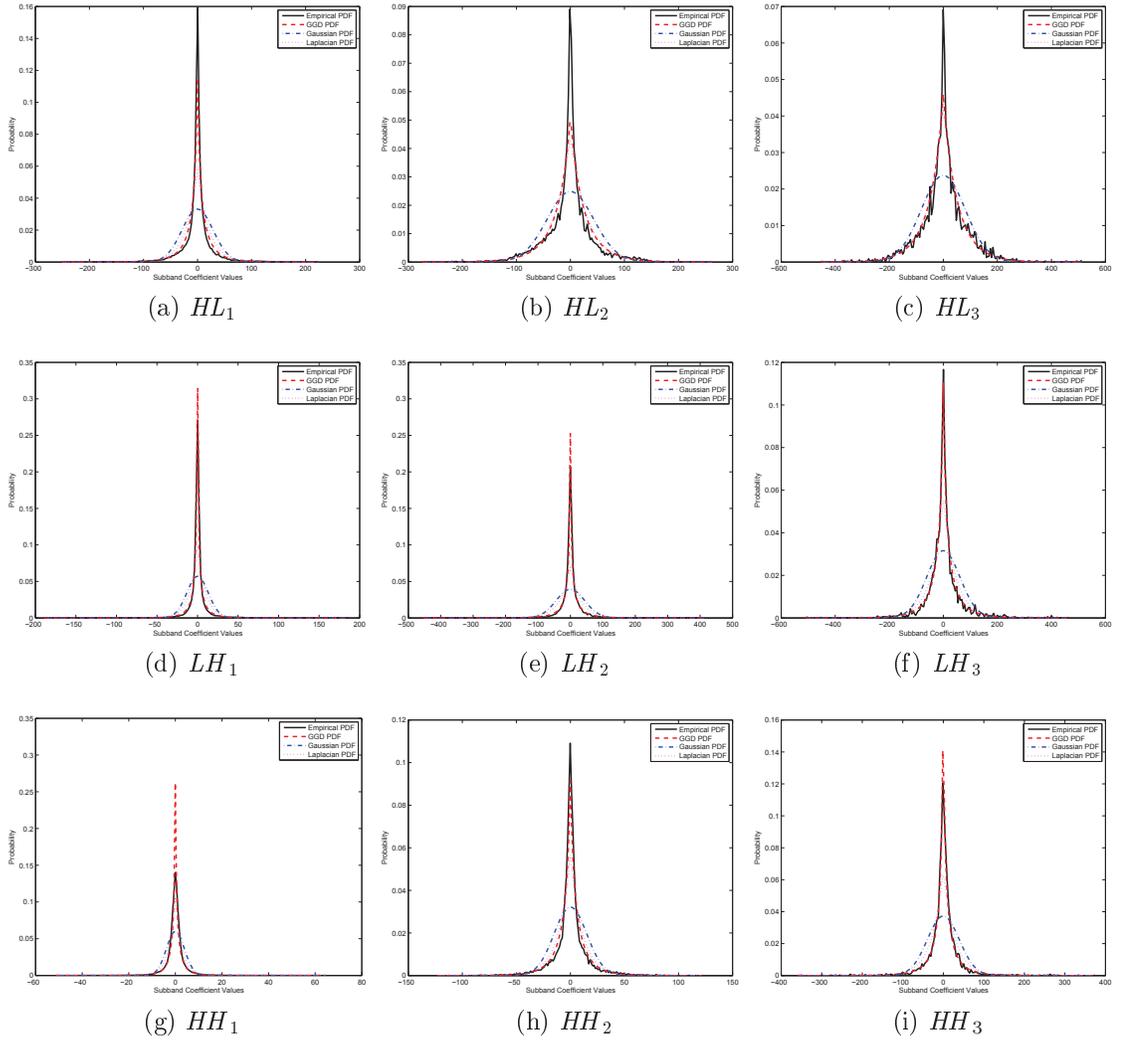


Figure 3.3: Modeling of the subband coefficients generated using a 3-level decomposition of frame 10 of the *Train* video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.

to model it. It can be expressed as:

$$KSD = \max_x \int [p_x(x) - \hat{p}_x(x)] dx. \quad (3.7)$$

- Kullback- Leibler Divergence (KLD): KLD is a non-symmetric measure of the difference between two probability distributions, viz., the empirical one and the one used to define it. It can be computed as:

$$KLD = \int \hat{p}_x(x) \ln \frac{\hat{p}_x(x)}{p_x(x)} dx. \quad (3.8)$$

Table 3.4 presents the computed values of KSD and KLD for the modeling of the empirical pdf corresponding to the subband coefficients for all the frames in the six videos used for evaluation using the GG, Gaussian and Laplacian distributions. The estimated values of these measures show that the GG distribution is able to fit the empirical data most precisely and have the least amount of distance and divergence from the empirical pdf. This high degree of matching between the empirical and the GG pdfs is because of the ability of the GG pdf to approximate for the sharp-peaked and heavy-tailed nature of the empirical pdf more accurately.

3.3 Spatio-Temporal Modeling of Subband Coefficients

The spatial modeling of the subband coefficients has been discussed in the previous section and is the one generally exploited when exploiting the various properties of the subband coefficients, such as sparseness, clustering and inter-subband correlations, for various applications. However, not much work has been done to

Test Image	KSD			KLD		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
	HH					
GG PDF	0.0976	0.1451	0.1553	0.1147	0.0936	0.1305
Laplacian PDF	0.2569	0.2965	0.2573	0.1420	0.2944	0.3982
Gaussian PDF	0.3540	0.3925	0.3473	0.4538	0.7367	0.8596
	LH					
GG PDF	0.1333	0.1159	0.1099	0.1002	0.1160	0.2064
Laplacian PDF	0.2717	0.2222	0.1635	0.3557	0.3968	0.3630
Gaussian PDF	0.3878	0.3337	0.2735	0.8053	0.8917	0.7327
	HL					
GG PDF	0.1420	0.1002	0.1251	0.1183	0.1203	0.1667
Laplacian PDF	0.2499	0.1824	0.1789	0.4705	0.3871	0.3877
Gaussian PDF	0.3459	0.2942	0.2789	0.9694	0.8266	0.7688

Table 3.4: Values of the metrics KSD and KLD for the modeling of the subband coefficients using GG, Gaussian and Laplacian distributions. The average values of KLD and KSD are obtained using all the frames of the 6 different test sequences.

utilize any relationship that exists between the subband coefficients in the temporal direction.

Figure 3.4 (a) shows two consecutive frames of the *Trevor* video sequence where the background is stationary while the object in the foreground (man) is moving. As is evident, there exists a lot of redundancy in the temporal direction between any two corresponding pixels at the same location in adjacent frames and this redundancy has been successfully exploited in applications like video coding. Even in sequences depicting motion, the temporal correlations persist in the regions not affected by motion. Motion primarily causes a change in the location of the edges from one frame to another. Since wavelets are able to preserve the edge information owing to spatial clustering, this motion, in the form of movement of edges, can also be captured and represented in the wavelet domain. The edges of the objects that are immobile are represented at the same locations in the wavelet domain representation of different frames, while the spatially uniform regions in

space are mapped to relatively low values in the wavelet domain. Thus, it is feasible to exploit the temporal redundancy of a video sequence in the wavelet domain as well.

One of the subbands (HL_1) corresponding to the 3-level wavelet decomposition of the two successive video frames is shown in Figure 3.4 (b). As can be seen, wavelets act as edge detectors and spatial clustering of significant coefficients along the spatial edges can be observed in the subband representation. Thus, the motion occurring in the video, in the form of movement of edges, can also be captured and represented in the wavelet domain and this concept is explored here.

Figure 3.4 (c) presents the difference of the HL_1 subbands corresponding to the two successive frames of the *Trevor* video sequence. The observed coefficients after the difference signify the change in the scene from one frame to the next which is caused due to motion. Thus, if \mathbf{X}^n is the set of subband coefficients for the current frame, n , and \mathbf{X}^{n-1} is the set of subband coefficients for the previous frame, $n - 1$, then \mathbf{X}^n can be represented as

$$\mathbf{X}^n = \mathbf{X}^{n-1} + \mathbf{M}^{n-1} \quad (3.9)$$

where \mathbf{M}^{n-1} is the motion noise introduced in the subband representation of the frame at instance $n - 1$ and can be explicitly defined as

$$\mathbf{M}^{n-1} = \{M_{ij}^{n-1} | i = 1, 2, \dots, L \text{ and } j = 1, 2, \dots, K\} \quad (3.10)$$

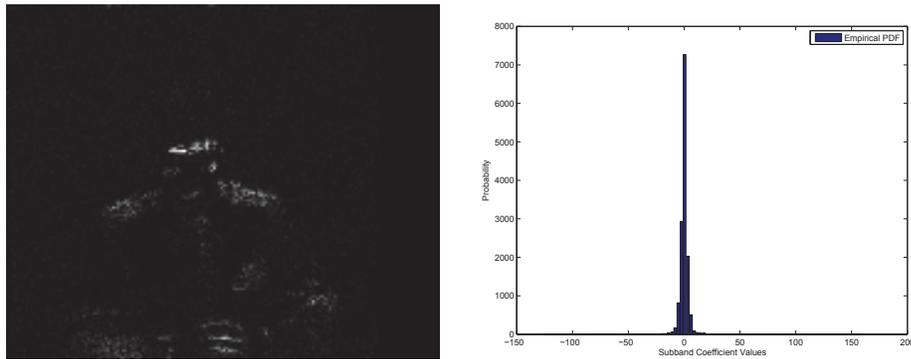
where, (i, j) represent the row and column coordinates of the wavelet coefficients, respectively, n denotes the transformed frame index, and L and K are the number of rows and columns in each wavelet domain representation, respectively. Since the observed motion noise is more pronounced around the edges and textured



(a) Original Frames 20 and 21 from *Trevor* Sequence



(b) HL_1 subbands



(c) Difference of the two subbands shown above and its histogram

Figure 3.4: (a) Two consecutive frames of *Trevor* video sequence in time domain; (b) HL_1 subband in the wavelet decomposition of two consecutive frames which are shown in (a). The subband coefficient values have been rescaled for visualization; and (c) The difference between the two subbands and the corresponding histogram.

areas, the wavelet coefficients corresponding to the motion noise also demonstrate the sparseness and clustering properties.

Figure 3.4 (c) also presents the histogram of the coefficients corresponding to one of the subbands of the motion noise frame wavelet coefficients. Similar to the histograms of the original frame subband coefficients, the histogram of the motion noise frame subband coefficients exhibits a sharp peak around its mean and is heavy-tailed. Thus, similar to the modeling of the coefficients belonging to the different subbands in \mathbf{X}^n , the coefficients belonging to the different subbands in \mathbf{M}^{n-1} (or \mathbf{M}^n corresponding to the time instant $n + 1$) can be fitted using the GGD prior as:

$$p_m(m) = \frac{\nu_m}{2s_m\Gamma(\frac{1}{\nu_m})} \exp(-|\frac{x}{s_m}|^{\nu_m}), \quad s_m, \nu_m > 0 \quad (3.11)$$

where $p_m(m)$ is the pdf of the random variable m , and the shape parameter, and ν_m and s_m are the model parameters. The variance and kurtosis of this random variable can be defined as

$$\sigma_m^2 = s_m^2 \frac{\Gamma(\frac{3}{\nu_m})}{\Gamma(\frac{1}{\nu_m})}, \quad (3.12)$$

and

$$\kappa_m = \frac{\Gamma(\frac{1}{\nu_m})\Gamma(\frac{5}{\nu_m})}{\Gamma^2(\frac{3}{\nu_m})}. \quad (3.13)$$

For each subband, the model parameters ν_m and s_m can be estimated by empirically estimating the variance and the kurtosis of the motion noise subband coefficients and solving (3.12) and (3.13) for the model parameters. The wavelet coefficients belonging to each subband of \mathbf{M}^{n-1} can be treated as i.i.d. random variables and can be obtained by subtracting the corresponding subband coefficients from successive frames.

To evaluate the modeling of the empirical distribution of the motion noise

Test Image	ν_m			s_m		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
Miss America	1.1045	0.6464	0.5743	3.1238	0.7307	0.6469
Hall	0.4985	0.4383	0.4561	0.4391	0.5059	0.7943
Train	0.3755	0.4263	0.3657	0.1469	0.7915	0.5362
Patrol Car	0.3514	0.4164	0.6166	0.0653	0.5596	11.0135
Salesman	0.8128	0.9840	1.2870	5.2188	14.3294	26.0670
Trevor	0.6823	0.4723	0.5393	0.9785	0.4398	1.4266

Table 3.5: Average value of the generalized Gaussian distribution parameters, s_m and ν_m for the modeling of the motion noise wavelet coefficients for some test videos for the LH subbands

Test Image	ν_m			s_m		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
Miss America	0.4534	0.3693	0.3650	0.1397	0.0956	0.1555
Hall	0.4577	0.4949	0.5679	0.2446	0.7312	1.2481
Train	0.5954	0.7241	0.6932	6.8438	11.8236	11.2135
Patrol Car	0.6882	0.6878	0.7264	8.8621	12.0397	22.2666
Salesman	1.3339	1.4174	1.7258	36.6167	77.8455	228.4577
Trevor	0.3922	0.4347	0.5218	0.1852	0.5918	2.0075

Table 3.6: Average value of the generalized Gaussian distribution parameters, s_m and ν_m for the modeling of the motion noise wavelet coefficients for some test videos for the HL subbands

Test Image	ν_m			s_m		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
Miss America	0.9578	0.6125	0.4988	2.7264	0.5230	0.3509
Hall	0.6707	0.5224	0.4668	0.4780	0.7774	0.5914
Train	0.4502	0.6665	0.5314	0.2637	3.3037	2.4755
Patrol Car	0.3141	0.3424	0.4162	0.0144	0.0818	0.7817
Salesman	0.7727	1.1352	1.1330	4.8446	16.6685	24.0621
Trevor	0.7770	0.4954	0.5349	0.9869	0.4440	1.1497

Table 3.7: Average value of the generalized Gaussian distribution parameters, s_m and ν_m for the modeling of the motion noise wavelet coefficients for some test videos for the HH subbands

wavelet coefficients by the GG distribution, six different video sequences have been used. All the frames in each of the video sequences have been transformed to the wavelet domain using a three level decomposition and the successive wavelet domain frames were subtracted from one another to obtain the motion noise coefficients. The resulting coefficients corresponding to each of the subbands were then approximated using the GG distribution. The model parameters (ν_m and s_m) for each subband were estimated from its variance and kurtosis as described above. Tables 3.5, 3.6 and 3.7 present the estimated values of the model parameters for each of the subbands for the six videos used for evaluation. The results are obtained by averaging the model parameters for corresponding subbands over all the frames of the respective video sequences. To compare the relative performance of this approximation, the same empirical subband data was also modeled using the Gaussian and Laplacian distributions. Figures 3.5 and 3.6 show the fitted histograms of all the subbands belonging to motion noise coefficients corresponding to the frames 20 and 21 of *Trevor* and frames 10 and 11 of *Train* video sequences, respectively. On visual inspection, the GGD prior is able to model the sharp peaks in the histograms around the zero mean as well as the heavy tails much more precisely than the Gaussian or the Laplacian distributions.

For a quantitative analysis of the fitting of the empirical pdf using the three *a priori* distributions, Table 3.8 presents the computed values of KSD and KLD for all the six video sequences used for evaluation. The presented values of these measures show that the GG distribution is able to approximate for the characteristics of the empirical pdf most accurately with the least amount of distance and divergence.

Thus, the subband coefficients corresponding to the individual frames in a video sequence have been successfully modeled in a unified spatio-temporal frame-

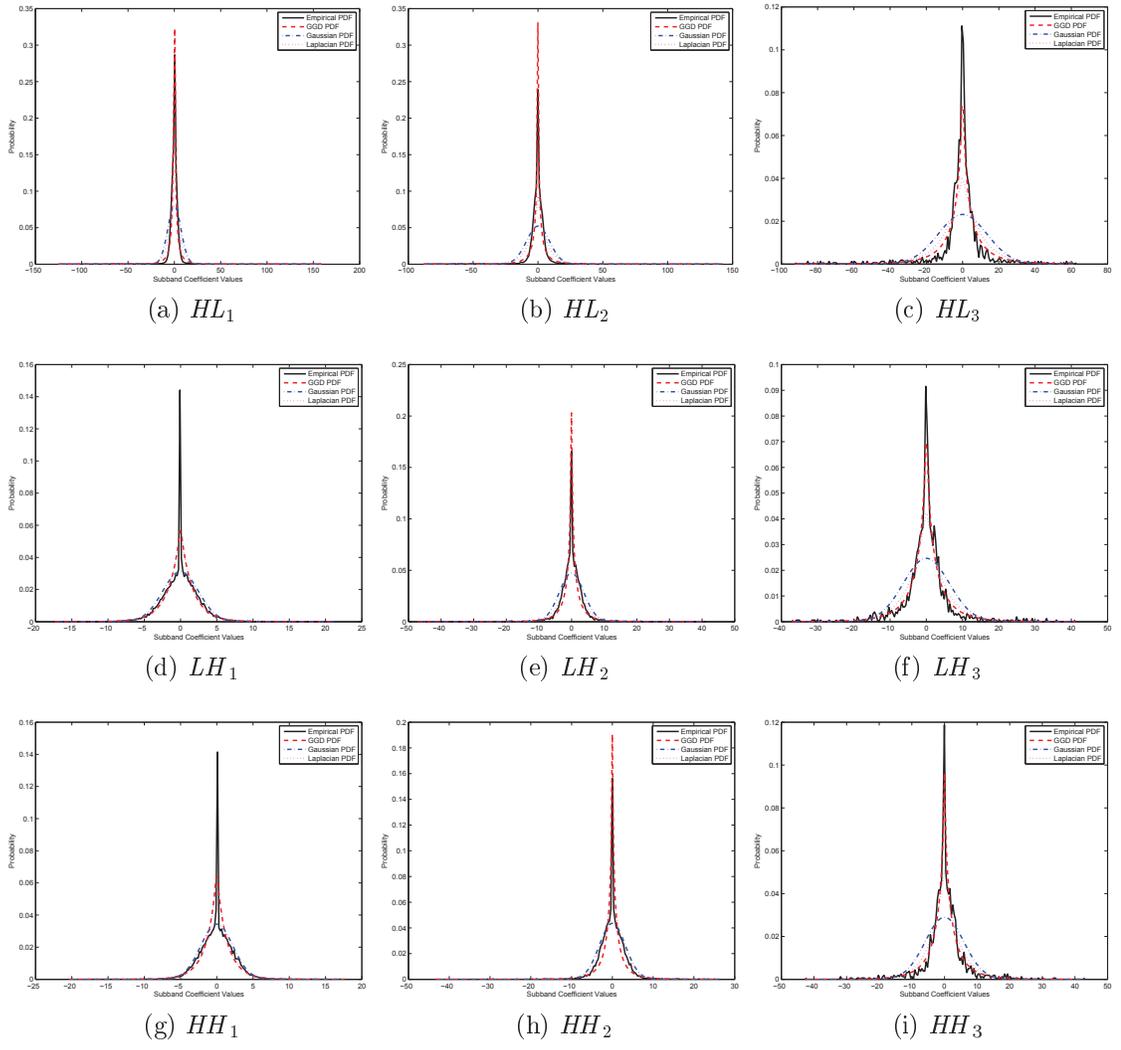


Figure 3.5: Modeling of the motion noise subband coefficients for frames 20 and 21 of the *Trevor* video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.

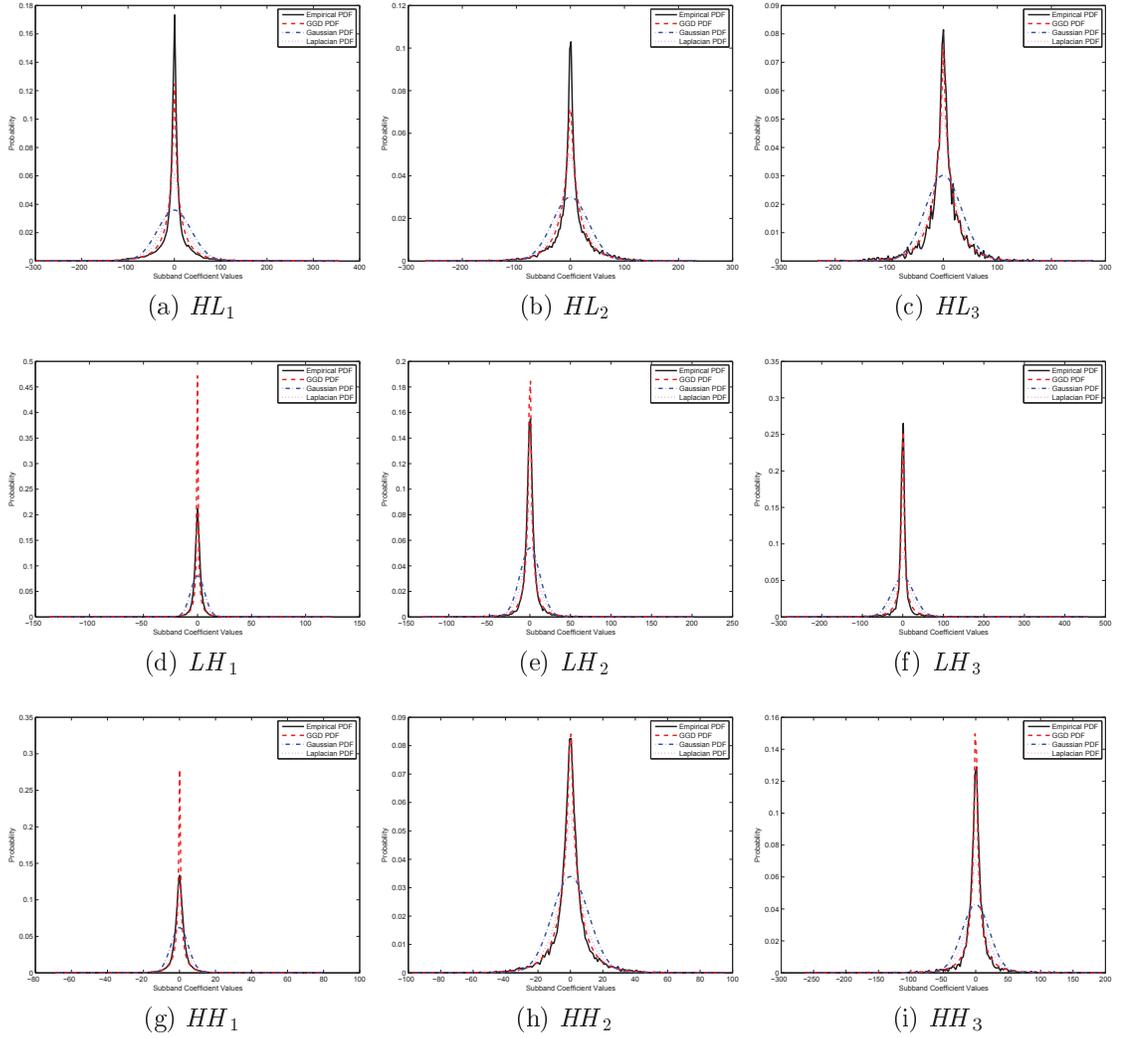


Figure 3.6: Modeling of the motion noise subband coefficients for frames 10 and 11 of the *Train* video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.

Test Image	KSD			KLD		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
	HH					
GG PDF	0.1166	0.1464	0.1746	0.0942	0.1187	0.1430
Laplacian PDF	0.2451	0.2715	0.2764	0.1020	0.1766	0.2761
Gaussian PDF	0.3416	0.3695	0.3606	0.3284	0.4693	0.6559
	LH					
GG PDF	0.1056	0.1403	0.1619	0.0819	0.1062	0.1784
Laplacian PDF	0.2537	0.2777	0.2570	0.1769	0.3201	0.4787
Gaussian PDF	0.3653	0.3730	0.3439	0.4824	0.7323	0.9369
	HL					
GG PDF	0.1167	0.1313	0.1650	0.1625	0.1473	0.1686
Laplacian PDF	0.2378	0.2428	0.2490	0.1818	0.1927	0.2494
Gaussian PDF	0.3342	0.3401	0.3269	0.4787	0.5001	0.5587

Table 3.8: Values of the metrics KSD and KLD for the modeling of the motion noise subband coefficients using GG, Gaussian and Laplacian distributions. The average values of KLD and KSD are obtained using all the frames of 6 different test sequences.

work using a single *a priori* distribution.

3.4 Conclusion

In this chapter, a unified spatio-temporal framework for the modeling of the video sequences in the wavelet domain has been proposed. The subband coefficients generated using the application of 2D DWT on the input video sequence frames have been modeled using the generalized Gaussian distribution. Via extensive results it has been shown that the GG prior used for the subband coefficients offers a very accurate and close fit to the empirical pdf exhibited by the subband coefficients.

It has been shown that similar to the temporal redundancies that exist in the time domain, these redundancies also exist in the corresponding subband coefficients belonging to successive frames. A modeling framework has been proposed

which models these temporal redundancies using the same GG prior that has been used to model the spatial subband coefficients. This modeling has been validated using extensive experimentation and the corresponding results are also presented.

In the following chapters, it will be shown that is possible to use this spatio-temporal framework for the modeling of video sequences in the wavelet domain in a variety of ways to develop noise reduction algorithms.

Chapter 4

Wavelet Domain Video Denoising Using Temporal DCT and Adaptive Thresholding

4.1 Introduction

As discussed in Chapter 1, the corruption of digital video sequences by noise is a common problem and this noise needs to be reduced before any display of the video sequences for visual purpose or before any further processing of the sequence for other applications such as segmentation and compression.

The existence of the spatio-temporal redundancies in the video sequences, in the time as well as wavelet domains, has been shown in Chapter 3. Despite this fact, while the spatial redundancies have been successfully exploited in the wavelet domain for the noise reduction in individual frames, the incorporation of the temporal redundancies in the noise reduction frame has been primarily in the time domain and *ad hoc* in nature (e.g., [29, 52, 59]). Thus, in this chapter, a fully

wavelet domain-based algorithm is proposed for video noise reduction, which is based on the spatio-temporal framework for video modeling discussed in Chapter 3.

The proposed algorithm requires only the local neighboring frames for its operation and before utilizing the spatial redundancies in the wavelet domain for noise reduction, the temporal subband coefficients are pre-processed using 1D discrete cosine transform (DCT) to reduce the redundancies in the temporal direction. It is shown that, even after the application of the DCT, the inter-subband spatial correlations still persist and that it is possible to model the decorrelated subband coefficients using the generalized Gaussian distribution . Based on this modeling, the resultant wavelet coefficients are effectively denoised using a low-complexity, hierarchically adapted thresholding scheme.

This chapter is organized as follows. A formulation of the noise reduction problem is provided in Section 4.2, while a description of the proposed system model is provided in Section 4.3. In Section 4.4, DCT is explored as a means for removing the temporal redundancy in the corresponding wavelet coefficients in consecutive frames. The hierarchically adapted threshold for the shrinkage of the noisy wavelet coefficients is discussed in Section 4.5. Experimental results comparing the proposed algorithm to several state-of-the-art techniques are presented in Section 4.6.

4.2 Problem Formulation

This chapter is concerned with noise reduction in digital video and concentrates on the reduction of independent and identically distributed (i.i.d.), additive white Gaussian noise (AWGN), which is the most prevalent and thus, the most com-

monly considered type of noise in literature. The noisy video signal, g_{ij}^n , which is discretized both in the spatial and temporal coordinates, can be modeled by

$$g_{ij}^n = f_{ij}^n + \eta_{ij}^n, \quad i = 1, \dots, L \text{ and } j = 1, \dots, K \quad (4.1)$$

where (i, j) denotes the spatial coordinates, n denotes the frame index, L and K are the number of rows and columns in each input frame, and f_{ij}^n and η_{ij}^n denote the original video signal and the AWGN, respectively. The bold-faced notation is used to represent the set of pixels in the n^{th} noisy and noise-free frames and the AWGN as \mathbf{g}^n , \mathbf{f}^n and η^n , respectively. The AWGN can be modeled as a normal random variable with zero mean and σ_η^2 variance:

$$\eta^n \in \mathcal{N}(0, \sigma_\eta^2) \quad (4.2)$$

The video sequence filtering problem is to find an estimate \hat{f}_{ij}^n of the true video signal f_{ij}^n based on the noisy observation g_{ij}^n .

Each of the input noisy frames, \mathbf{g}^n , is transformed to the wavelet domain using a critically-sampled (i.e., non-redundant) discrete wavelet transform (DWT)

$$\mathbf{Y}^n = DWT(\mathbf{g}^n) \quad (4.3)$$

where \mathbf{Y}^n is the set of the noisy wavelet coefficients for the n^{th} frame. Owing to the linearity and the orthogonality of the DWT operation, the corrupting noise in the wavelet domain is also AWGN. In the wavelet domain, the video signal can be represented as

$$Y_{ij}^n = X_{ij}^n + \epsilon_{ij}^n, \quad i = 1, \dots, L \text{ and } j = 1, \dots, K \quad (4.4)$$

where (i, j) and n represent the spatial coordinates and the frame index, respectively, Y_{ij}^n and X_{ij}^n are respectively the noisy and noise-free wavelet coefficients, and ϵ_{ij}^n are the independent and identically distributed (i.i.d.) samples from an AWGN process having a zero mean and variance σ_ϵ^2 (where $\sigma_\epsilon^2 = \sigma_\eta^2$).

Thus, in the wavelet domain, the noise reduction problem is to find an estimate $\hat{\mathbf{X}}^n$ of the original frame \mathbf{X}^n based on the noisy observation \mathbf{Y}^n .

4.3 System Description

There exists spatial as well as temporal redundancy or correlation in the wavelet domain representation of the input video sequence. The present approach tries to minimize the temporal redundancy of the wavelet coefficients in the current frame from those in its immediate neighborhood before any spatial noise reduction is performed. This is achieved via the application of the DCT along the temporal direction and serves to improve the efficiency of the algorithm and reduce noise more aggressively. In this work, the application of the 2D DWT followed by DCT in the temporal direction is preferred over the 3D DWT, as the latter processes the video sequence as a 3D block of data and thus, is computationally very expensive, has large memory requirements and is not real-time. To restrict the memory requirements and computational burden to a minimum, only the frames in the immediate neighborhood of the current frame are utilized. The temporal redundancy in the wavelet domain representations of the noisy current, past and future frames is minimized as

$$\begin{bmatrix} D_{ij}^{n-\frac{N-1}{2}} \\ \vdots \\ D_{ij}^{n+\frac{N-1}{2}} \end{bmatrix} = DCT \begin{bmatrix} Y_{ij}^{n-\frac{N-1}{2}} \\ \vdots \\ Y_{ij}^{n+\frac{N-1}{2}} \end{bmatrix} \quad (4.5)$$

where \mathbf{D}^i ($n - \frac{N-1}{2} \leq i \leq n + \frac{N-1}{2}$) are the current, past and future frames after the application of the DCT. The resultant frames are then spatially denoised using wavelet thresholding as follows

$$\begin{bmatrix} \hat{\mathbf{D}}^{n-\frac{N-1}{2}} \\ \vdots \\ \hat{\mathbf{D}}^{n+\frac{N-1}{2}} \end{bmatrix} = \text{Denoising} \begin{bmatrix} \mathbf{D}^{n-\frac{N-1}{2}} \\ \vdots \\ \mathbf{D}^{n+\frac{N-1}{2}} \end{bmatrix} \quad (4.6)$$

where $\hat{\mathbf{D}}^i$ ($n - \frac{N-1}{2} \leq i \leq n + \frac{N-1}{2}$) are the denoised versions of the DCT transformed frames. After spatial noise reduction, the frames are operated upon by the inverse of the DCT operation as

$$\begin{bmatrix} \hat{X}_{ij}^{n-\frac{N-1}{2}} \\ \vdots \\ \hat{X}_{ij}^{n+\frac{N-1}{2}} \end{bmatrix} = \text{IDCT} \begin{bmatrix} \hat{D}_{ij}^{n-\frac{N-1}{2}} \\ \vdots \\ \hat{D}_{ij}^{n+\frac{N-1}{2}} \end{bmatrix} \quad (4.7)$$

where $\hat{\mathbf{X}}^n$ is the denoised estimate of the set of noise-free wavelet coefficients in the current frame. The output of the inverse discrete wavelet transform operation (IDWT) on the denoised coefficients in the current frame yields the denoised time domain pixel values as follows:

$$\hat{\mathbf{f}}^n = \text{IDWT}(\hat{\mathbf{X}}^n) \quad (4.8)$$

where $\hat{\mathbf{f}}^n$ is the required denoised estimate of the noise-free n^{th} frame in the time domain.

The block diagram of the proposed scheme is presented in Fig. 4.1.

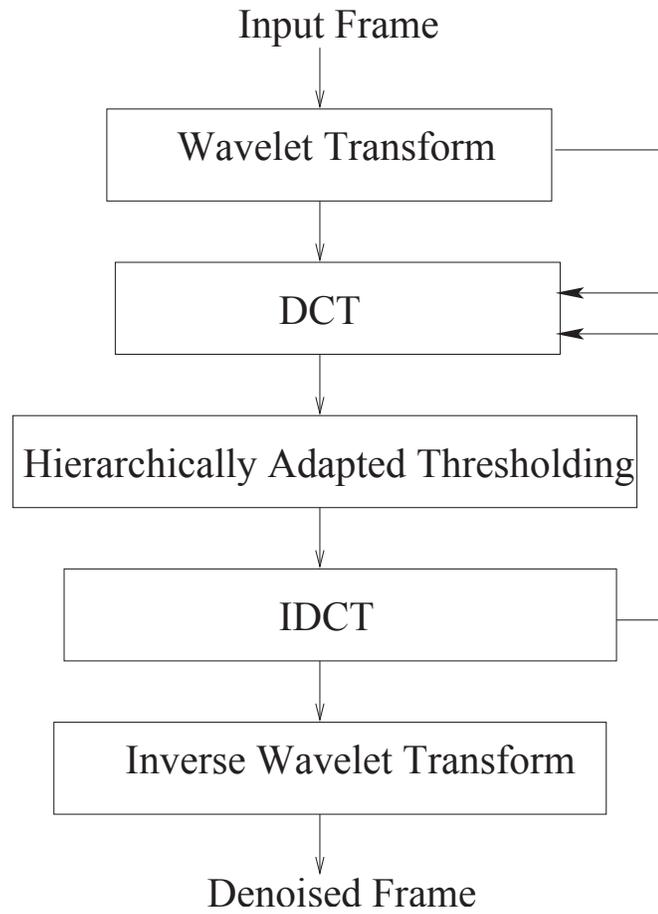


Figure 4.1: Block Diagram of the proposed video noise reduction scheme.

4.4 Temporal DCT

To process the corresponding subband coefficients in the temporal direction, a N -point DCT of the corresponding noisy wavelet coefficients in the $\frac{(N-1)}{2}$ past frames, $\frac{(N-1)}{2}$ future frames and the current frame n is used. This transform serves to reduce the temporal redundancy prior to the spatial denoising of the noisy subband coefficients. For a N -point DCT, the operations in (4.5) can thus be expressed as

$$D_{ij}^k = \alpha(k) \sum_{l=n-\frac{(N-1)}{2}}^{l=n+\frac{(N-1)}{2}} Y_{ij}^l \cos\left(\frac{\pi}{2N}(2(l-n+\frac{N-1}{2})+1)(k-n+\frac{N-1}{2})\right), \quad (4.9)$$

$$n - \frac{N-1}{2} \leq k \leq n + \frac{N-1}{2}$$

where

$$\alpha(0) = \sqrt{\frac{1}{N}}, \quad \alpha(k) = \sqrt{\frac{2}{N}} \quad \text{for } n - \frac{N-1}{2} + 1 \leq k \leq n + \frac{N-1}{2} \quad (4.10)$$

The inverse transform can be expressed as

$$\hat{X}_{ij}^l = \sum_{k=n-\frac{(N-1)}{2}}^{k=n+\frac{(N-1)}{2}} \alpha(k) \hat{D}_{ij}^k \cos\left(\frac{\pi}{2N}(2(l-n+\frac{N-1}{2})+1)(k-n+\frac{N-1}{2})\right), \quad (4.11)$$

$$n - \frac{N-1}{2} \leq l \leq n + \frac{N-1}{2}$$

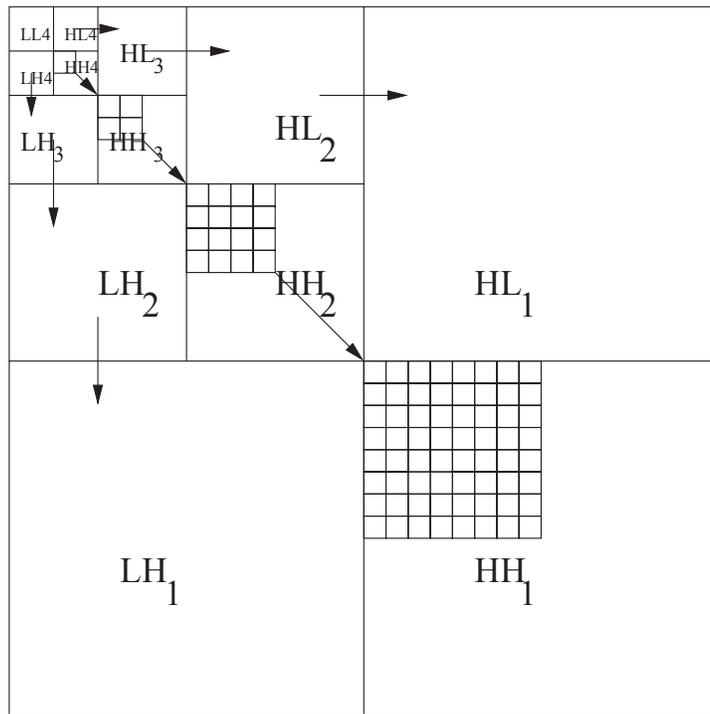
In order to keep the computational complexity low and to reduce the memory requirements, in this work we have process the current noisy frame using: (a) only the single past frame (2 total frames); (b) only one past and one future frame (3 total frames); and (c) two past and two future frames (5 total frames).

4.5 Hierarchically Adapted Wavelet Thresholding

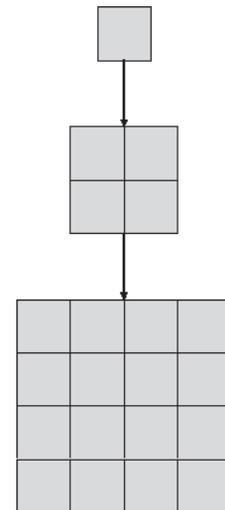
The critically-sampled DWT has a subband structure as shown in the Fig. 4.2. The subbands LH_k , HL_k and HH_k are called the *detail* subbands, where k is the scale varying from $1, \dots, J$ and J is the total number of decomposition levels. The LL_J subband is the low-resolution residue and the size of a subband at scale k is $\frac{L}{2^k} \times \frac{K}{2^k}$. For subbands LH_k, HL_k and HH_k , with scale $k \leq J - 1$, the parent subbands are defined as LH_{k+1}, HL_{k+1} and HH_{k+1} , respectively (see also Fig. 4). Similarly, for the subbands LH_k, HL_k and HH_k , $2 \leq k \leq J$, the children subbands are defined as LH_{k-1}, HL_{k-1} and HH_{k-1} , respectively. For a subband coefficient, D_{ij}^n , at the position (i, j) in a subband of frame n , the parent coefficient, P_{ij}^n , is the $(\lceil \frac{i}{2} \rceil, \lceil \frac{j}{2} \rceil)$ coefficient in the parent subband, where $\lceil \cdot \rceil$ maps the input value to the nearest lower integer.

The correlation that exists among the subband coefficients and their corresponding children has long been recognized (e.g., [51]). Consider \mathbf{Z}^n as the set of correlated subband coefficients that belong to a tree (as depicted in Fig. 4.2 (b)) in the wavelet representation of frame n . For this discussion, these wavelet coefficients belonging to a tree are treated as random variables and their correlation coefficient is defined as γ_Z^n . In the case where only the past frame is utilized for denoising, after the application of the DCT (see (4.9)), the resultant subband coefficients are proportional to the sum and difference of the subband coefficients of the input frames n and $n - 1$, respectively. If we denote by \mathbf{B} , the set of coefficients that are proportional to the sum of the two subband coefficients, and by \mathbf{C} , the set of coefficients that are proportional to the difference of the two subband coefficients, then the resultant coefficients would have a correlation given by

$$\gamma_B = \gamma_Z^{n-1} + \gamma_Z^n + 2\gamma_{cross} \quad (4.12)$$



(a) Subband structure



(b) Parent-child tree

Figure 4.2: Subband structure and parent-child relationship in a 4-level orthogonal discrete wavelet transform

$$\gamma_C = \gamma_Z^{n-1} + \gamma_Z^n - 2\gamma_{cross} \quad (4.13)$$

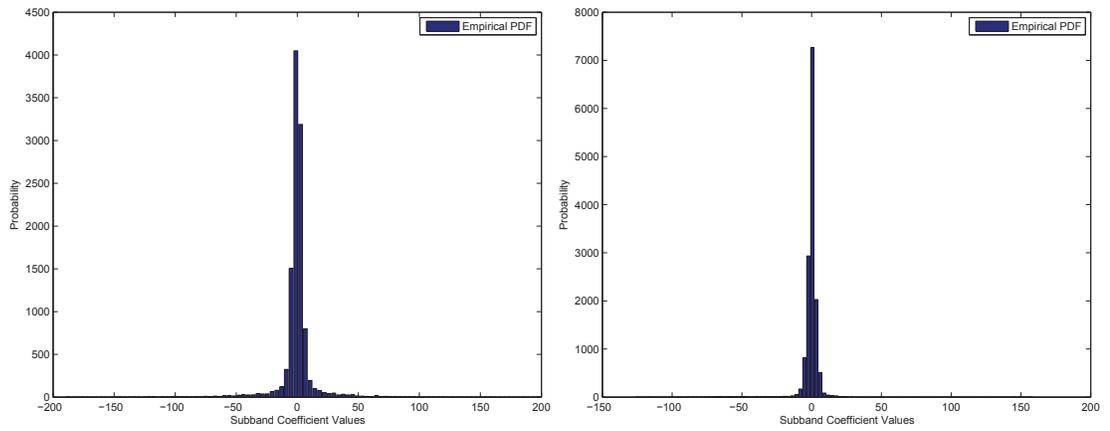
where γ_{cross} is the cross-correlation term. If the two input frames are exactly the same (representing no motion), then $\gamma_{cross} = \gamma_Z^{n-1} = \gamma_Z^n$. Thus, coefficients belonging to the sets **B** and **C** would also be correlated and would have the same parent-child correspondence as existed before the application of the DCT in the temporal direction. Only in the case when the two input frames are exactly the same, the difference coefficients, **C**, would all become zero. This reasoning is valid for the cases where more than two frames are used for denoising the current noisy frame.

Figure 4.3 (a) shows the one horizontal subband from the multi-resolution subband structures generated after a 4-level non-redundant orthogonal DWT and temporal DCT using (4.9). Even after the DCT operation to minimize temporal redundancies, the spatial clustering property of the wavelet transform is visible. As discussed above, there also exists a strong correlation among the corresponding coefficients in subsequent subbands. This correlation is depicted as in Fig. 4.2, where the tail of an arrow denotes the parent coefficients and the arrow head denotes the corresponding children coefficients, and is also evident in the two subsequent subbands shown from one decorrelated frame of *Miss America* sequence in Figure 4.4.

Figure 4.3 (b) presents the histograms of the original subband coefficients from the decorrelated wavelet coefficients from two successive frames. Noticing the evident sharp-peaked and heavy-tailed nature of these histograms, the coefficients belonging to the different subbands in the decorrelated frames **D**ⁿ have been fitted



(a) Decorrelated subsequent subbands



(b) Histograms of the decorrelated subsequent subbands

Figure 4.3: (a) Horizontal subbands from the wavelet and DCT transformed consecutive frames from the *Trevor* sequence; and (b) The corresponding histograms of the decorrelated subbands in (a)

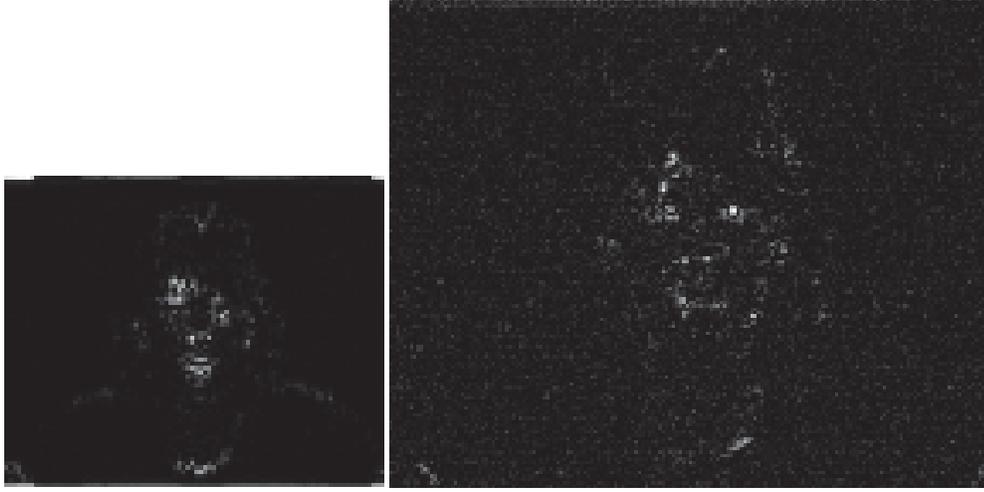


Figure 4.4: Subbands HL_2 and HL_1 of a frame from sequence Miss America after application of DCT.

using the GGD prior as:

$$p_d(d) = \frac{\nu_d}{2s_d\Gamma(\frac{1}{\nu_d})} \exp(-|\frac{x}{s_d}|^{\nu_d}), \quad s_d, \nu_d > 0 \quad (4.14)$$

where $p_d(d)$ is the pdf of the random variable d , and the shape parameter, and ν_d and s_d are the model parameters. The variance and kurtosis of this random variable can be defined as

$$\sigma_d^2 = s_d^2 \frac{\Gamma(\frac{3}{\nu_d})}{\Gamma(\frac{1}{\nu_d})}, \quad (4.15)$$

and

$$\kappa_d = \frac{\Gamma(\frac{1}{\nu_d})\Gamma(\frac{5}{\nu_d})}{\Gamma^2(\frac{3}{\nu_d})}. \quad (4.16)$$

For each subband, the model parameters ν_d and s_d can be estimated by using the empirically estimating the variance and kurtosis of the decorrelated subbands and solving 4.15 and 4.16 for the model parameters.

Six different video sequences are used to evaluate the modeling performance of the GG prior for the decorrelated subband coefficients. The model parameters

Test Image	ν_d			s_d		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
Miss America	0.5433	0.4863	0.6065	0.6868	1.1662	9.6918
Hall	0.4316	0.5967	0.7340	1.1104	13.6666	53.4711
Train	0.4447	0.4465	0.6266	1.2644	4.3889	27.1056
Patrol Car	0.3828	0.3976	0.6129	0.2454	0.6731	21.7615
Salesman	0.7773	0.9977	1.3071	4.8334	14.3425	27.5562
Trevor	0.7244	0.7356	0.8355	4.4302	11.6160	41.8386

Table 4.1: Average value of the generalized Gaussian distribution parameters, s_d and ν_d for the modeling of the decorrelated wavelet coefficients for some test videos for the LH subbands

Test Image	ν_d			s_d		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
Miss America	0.3098	0.4164	0.4719	0.0478	0.7457	4.1787
Hall	0.3880	0.5309	0.6086	0.6389	8.3130	32.6814
Train	0.6039	0.8475	0.9184	9.1894	41.7852	99.6454
Patrol Car	0.6940	0.7259	0.7844	10.9083	35.6565	100.9352
Salesman	1.3992	1.4113	1.3254	37.9322	79.3889	136.5275
Trevor	0.4118	0.4580	0.5482	0.3749	1.9296	10.7243

Table 4.2: Average value of the generalized Gaussian distribution parameters, s_d and ν_d for the modeling of the decorrelated wavelet coefficients for some test videos for the HL subbands

Test Image	ν_d			s_d		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
Miss America	0.7536	0.4122	0.5300	0.6263	0.2165	2.8438
Hall	0.3489	0.3866	0.5084	0.0529	0.5746	6.2386
Train	0.4915	0.6506	0.4928	0.5021	7.1599	6.4249
Patrol Car	0.3378	0.3543	0.3836	0.0309	0.1214	1.0528
Salesman	0.7415	1.0733	1.3464	4.2549	16.0229	26.4280
Trevor	0.5243	0.4544	0.5617	0.3899	0.7760	6.1517

Table 4.3: Average value of the generalized Gaussian distribution parameters, s_d and ν_d for the modeling of the decorrelated wavelet coefficients for some test videos for the HH subbands

(ν_d and s_d) for each subband are estimated from its variance and kurtosis as described above. Tables 4.1, 4.2 and 4.3 present the averaged estimated values of the model parameters for each of the subbands for the six videos used for evaluation. The same empirical subband data is also modeled using the Gaussian and Laplacian distributions to compare the modeling performance of the GG distribution. Figure 4.5 and 4.6 show the fitted histograms of all the subbands belonging to decorrelated coefficients corresponding to the *Trevor* and *Train* video sequences. On visual inspection, the GGD prior is able to model the sharp peaks in the histograms around the zero mean as well as the heavy tails much more precisely than the Gaussian or the Laplacian distributions.

Table 4.4 presents the computed values of KSD and KLD for all the six video sequences used for quantitative evaluation. The presented values of these measures show that the GG distribution has the least distance and divergence while fitting the empirical pdf.

With the assumption of a GGD prior for the decorrelated subband coefficients, an approximate MSE-optimal soft-threshold, T_{bayes} , has been proposed by Chang *et al.* [49]

$$T_{bayes} = \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_d} \quad (4.17)$$

where the estimated noise variance, $\hat{\sigma}_\epsilon^2$, is computed using a robust estimator [45] such as

$$\hat{\sigma}_\epsilon^2 = \left(\frac{\text{median}(|HH_1|)}{0.6745} \right)^2. \quad (4.18)$$

The signal power, $\hat{\sigma}_X$, is estimated by a truncated moment estimator

$$\hat{\sigma}_d = \sqrt{\max(0, \frac{1}{b} \sum_{ij} D_{ij}^n - \hat{\sigma}_\epsilon^2)}, \quad (4.19)$$

where b is the total number of coefficients in the considered subband.

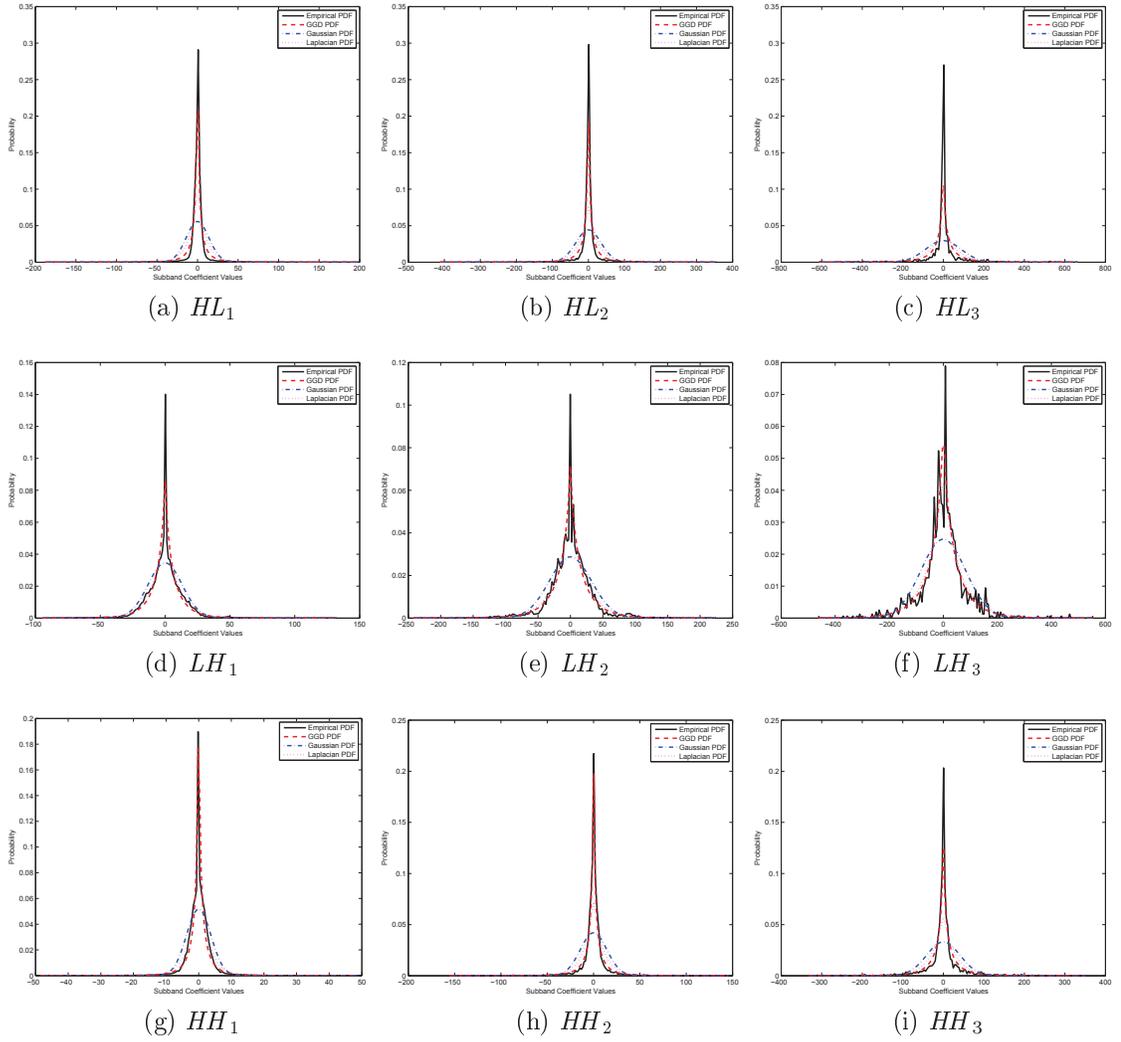


Figure 4.5: Modeling of the decorrelated subband coefficients for from the *Trevor* video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.

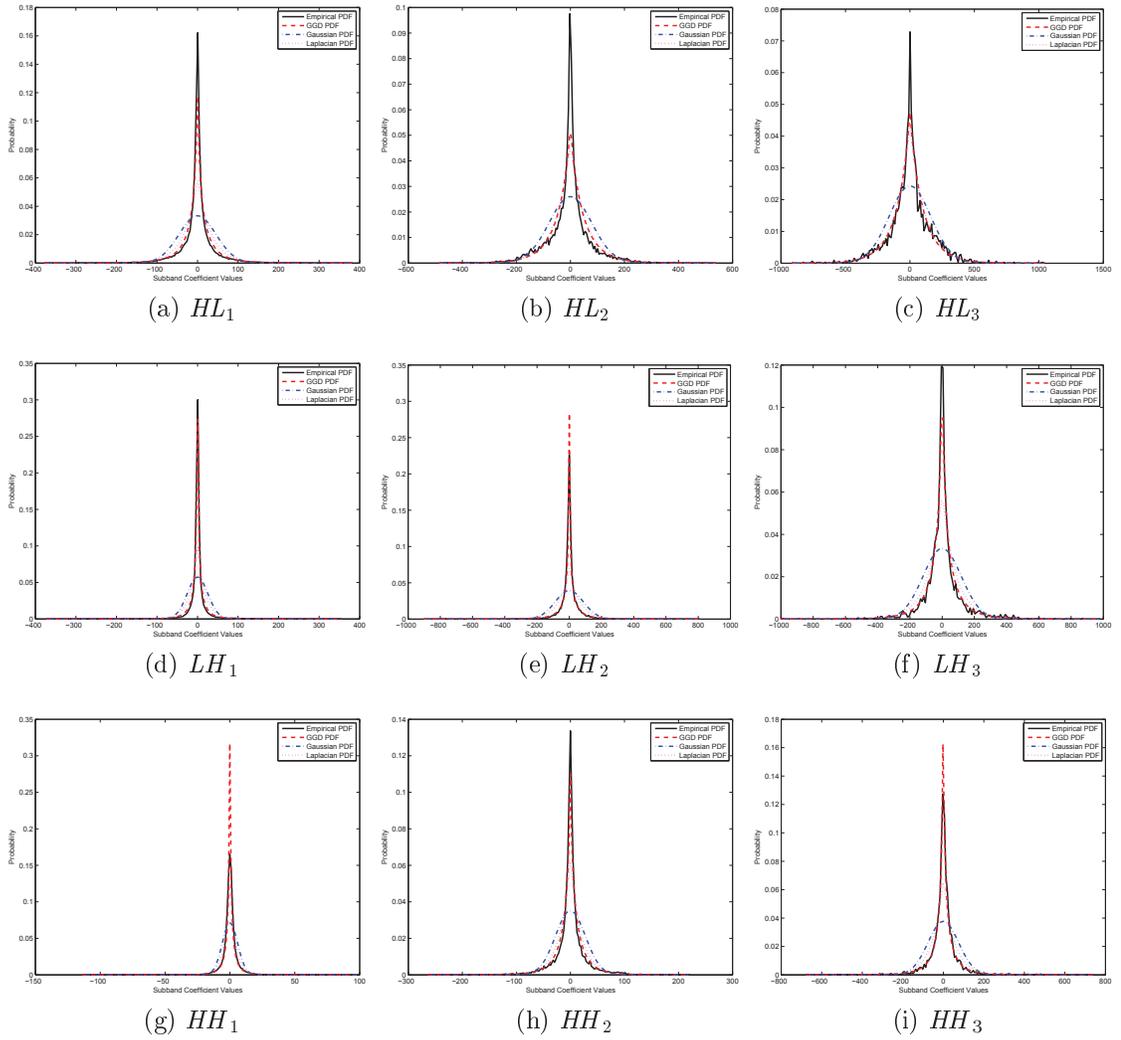


Figure 4.6: Modeling of the decorrelated subband coefficients from the *Train* video sequence. The empirical subband histogram has been fitted using the generalized Gaussian, Gaussian and Laplacian distributions.

Test Image	KSD			KLD		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
	HH					
GG PDF	0.1235	0.1442	0.1509	0.1049	0.0833	0.1388
Laplacian PDF	0.3024	0.2947	0.2543	0.1644	0.3396	0.4290
Gaussian PDF	0.3955	0.3845	0.3386	0.4734	0.7998	0.8956
	LH					
GG PDF	0.1378	0.1202	0.1117	0.1182	0.1202	0.2037
Laplacian PDF	0.2864	0.2180	0.1655	0.3978	0.2180	0.3687
Gaussian PDF	0.4037	0.3327	0.2682	0.8202	0.3327	0.7347
	HL					
GG PDF	0.1356	0.1042	0.1305	0.1162	0.1328	0.1762
Laplacian PDF	0.2556	0.1859	0.1850	0.5349	0.4354	0.4109
Gaussian PDF	0.3496	0.3004	0.2893	1.0371	0.8809	0.7963

Table 4.4: Values of the metrics KSD and KLD for the modeling of the decorrelated subband coefficients using GG, Gaussian and Laplacian distributions. The average values of KLD and KSD are obtained using all the frames of 6 different test sequences.

Although the above threshold is nearly subband optimal, it does not consider the properties of the individual coefficients in the subband. This threshold has been better adapted to obtain better denoising results [81]. Since the coefficient values in the subsequent subbands are correlated, a coefficient having a high value in a parent subband, which usually represents an edge or high frequency information, gives a fair indication that the child coefficient would more probably be a relatively high value than not. A zero in a parent coefficient indicates a zero or low value in the child coefficient. Using this philosophy, we scale the nearly MSE-optimal threshold computed for the entire subband using the individual parent coefficients. This yields an adaptive threshold which acts individually on each subband coefficient. We do the computations for the subbands in decreasing order of scales and thus, denoise the parent coefficient earlier than the progeny. We use the denoised parent coefficients to get an accurate estimate of the correspond-

ing subband coefficient values in the actual image sequence. The hierarchically adapted threshold (HAT) obtained can be expressed as

$$HAT_{ij}^n = \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_d \left(\alpha + \beta \frac{|P_{ij}^n|}{\max_{ij}(|P_{ij}^n|)} \right)} \quad (4.20)$$

where P_{ij} are the soft-thresholded parent coefficients. The variable in the denominator, $\alpha > 0$, has been added to avoid the singularity that occurs when the thresholded parent coefficient is zero. The variable $\beta > 0$ is used to vary the contribution of the parent coefficient in the final scaling of the subband optimal threshold. The parameters α and β are experimentally determined and their estimation is discussed in Section 4.6.2.

In case the relative value of the parent coefficient is low (e.g., zero), the threshold value is made higher to threshold the current subband coefficient more strongly as the value of the original coefficient would most likely be low and the rest of the contribution to this coefficient is due to white noise. In case of a high relative value of a parent coefficient, the threshold is lowered so as to retain more of the current coefficient. The threshold adapts to the level of noise in the signal. In the case when the noise level is low, a low magnitude threshold preserves most of the signal while eliminating noise. On the other hand, when noise is predominant, a high threshold value removes most of the noise from the signal and retains the significant coefficients corresponding to the image spatial features in the image.

All the subbands, with the exception of the LL_J subband, are thresholded using (4.20). This procedure is performed with both the decorrelated frames. The denoised estimate, \hat{D}_{ij}^n , using the noisy decorrelated frame in the wavelet

domain, D_{ij}^n , is obtained using soft thresholding [46] as

$$\hat{D}_{ij}^n = \text{sign}(D_{ij}^n)(\max(0, |D_{ij}^n| - HAT_{ij}^n)). \quad (4.21)$$

The two denoised and decorrelated frames obtained after soft thresholding are operated upon by the inverse of the DCT operation using (4.11). $\hat{\mathbf{X}}^n$ obtained as a result of this operation is the denoised estimate of the set of original noise-free wavelet coefficients in the frame n , \mathbf{X}^n .

The final denoised estimate of the original pixel set is obtained using (4.8).

4.6 Experimental Results and Discussion

4.6.1 Performance Criteria

The following criteria are used to evaluate the performance of different noise reduction algorithms:

1. **Objective criteria:** The peak signal to noise ratio (PSNR) is chosen as the objective criterion for the evaluation of noise reduction effectiveness. It is one of the most popular objective criteria, along with mean squared error (MSE) and signal to noise ratio (SNR). Given the n^{th} frame in an original sequence and in a corrupted or estimated version of the original, the PSNR can be computed as:

$$\text{PSNR}^n = 10 \log_{10} \frac{255^2}{\frac{1}{KL} \sum_{i=1}^K \sum_{j=1}^L (f^n(i, j) - \hat{f}^n(i, j))^2} \text{ dB} \quad (4.22)$$

where K and L are the number of columns and the number of rows in one frame of the video sequence (one field in case of interlaced video), respec-

tively.

2. **Subjective criteria:** The performance of the noise reduction algorithms is also compared based on the subjective evaluation of the video sequences. The commonly used objective criteria, like PSNR, MSE and SNR, although mathematically tractable, do not conform closely to the human visual system. This makes the subjective evaluation of the filtered sequences necessary.

4.6.2 Experimental Results

To validate the efficacy of the proposed video denoising scheme without motion compensation (DCWT) we have compared its performance with some state-of-the-art noise reduction algorithms, such as the 3D Rational filter (RF) [26], 3D KNN filter [24], 3D alpha-trimmed mean (ATM) filter [24] and the multiple class averaging (MCA) algorithm [29] (which operates in the wavelet domain). For the evaluation of the proposed scheme wherein motion in the previous (and future) frames is compensated (MC-DCWT), its performance is compared with the 3D LMMSE filter [35], the adapted weighted averaging (AWA) filter [35], the joint Kalman and Wiener filter (JKW) [39] and the MC3DWTr [58] (wavelet domain). To evaluate the effect of the application of the DCT in the temporal direction on the overall denoising of the video sequence, the output generated using only the spatial denoising algorithm (HAT) has also been used for comparison. For comparison with KNN and ATF filters, a window size that yields the maximum average output PSNR over the 60 frames is selected in the final algorithm. This results in a window size of $3 \times 5 \times 5$ for both these filters. The use of three temporal frames helps maintain the stationary nature of the pixels in the window, while the 5×5 size spatial window removes noise by averaging while keeping blurring to an

acceptable level.

We use three sequences in the CIF format, *Hall* (Ha), *Foreman* (Fo) and *Miss America* (MA), and two interlaced sequences in PAL format, *Patrol Car* (PC) and *Train* (Tr). Peak signal to noise ratio (PSNR) and visual evaluation are used to evaluate the relative performance of the various algorithms. The video sequences are corrupted with three different additive white Gaussian noise powers, so that the input PSNR to the different filtering schemes is 20 dB, 25 dB and 30 dB, respectively.

To estimate the optimal values of the parameters α and β , the PSNR values for a dataset of images corrupted with ten different noise powers are calculated over a range of values α and β . The values of α and β corresponding to the maximum of the averaged PSNR over the entire dataset are used to generate the final results. These estimated values of the parameters, corresponding to the maximum average output PSNR, are $\alpha = 0.43$ and $\beta = 4.3$.

Table 4.5 presents the average PSNR values for the different non-motion-compensated filter outputs with the sequences having PSNRs of 20 dB, 25 dB and 30 dB as input. We also compare the performance of the proposed spatio-temporal filter to that of only the spatial filter in these tables. The proposed filter, DCWT, yields an average PSNR gain in the range of 6 to 14 dB over the corrupted sequence and gains of 0.8-7 dB over the compared schemes. For sequences having high motion content, such as Tr or PC, the amount of PSNR improvement is less than that for sequences having lower motion content (e.g., MA and Ha). This is supported by the fact that the PSNR improvement of the spatio-temporal filter over the spatial filter (HAT) decreases with an increase in the motion content. While for sequences having a low motion content, the proposed filter yields an improvement of 1-1.5 dB over HAT, for sequences with high motion content,

		Non-motion-compensated Filters					Motion-compensated					Spatial Filter
		ATM	KNN	RF	MCA	DCWT	LMMSE	AWA	JKW	MC3DWTr	MC-DCWT	(HAT)
Input PSNR = 20 dB												
MA	30.37	29.0	29.29	31.21	34.83	30.96	29.01	30.65	35.35	34.98	33.6	
Ha	27.66	27.29	27.31	28.09	29.11	27.35	27.75	28.68	29.57	29.35	27.58	
Fo	27.96	27.75	28.11	28.07	29.54	26.14	28.09	27.64	29.85	29.77	27.34	
PC	22.94	24.14	25.97	26.15	27.29	25.5	25.6	24.89	27.99	27.85	26.84	
Tr	22.38	23.89	24.27	25.53	25.88	25.3	25.22	25.59	26.72	26.43	25.31	
Average	26.25	26.41	26.95	27.81	29.33	27.05	27.15	27.48	29.89	29.67	28.13	
Input PSNR = 25 dB												
MA	33.61	33.45	34.18	34.92	36.78	34.57	33.36	32.22	37.32	36.93	35.82	
Ha	28.74	29.98	30.34	32.04	32.47	30.86	31.07	32.37	32.92	32.71	30.99	
Fo	30.58	30.37	31.66	31.44	32.15	30.13	29.85	29.42	32.59	32.44	30.75	
PC	23.06	25.18	27.93	29.95	30.66	28.83	29.08	27.41	31.31	31.12	30.23	
Tr	22.9	24.8	25.43	29.25	29.43	28.74	28.6	28.25	30.08	29.96	29.02	
Average	27.78	28.75	29.91	31.52	32.30	30.63	30.39	29.97	32.84	32.63	31.36	
Input PSNR = 30 dB												
MA	37.69	37.19	37.53	37.78	38.72	37.66	37.3	37.25	38.97	38.89	38.13	
Ha	29.2	31.38	31.64	35.13	35.61	34.66	34.97	35.11	36.05	35.92	34.39	
Fo	31.25	32.00	33.44	34.65	34.96	33.95	31.15	31.72	35.41	35.26	33.82	
PC	23.06	25.48	28.62	33.2	34.27	32.38	32.55	29.85	34.98	34.85	34.12	
Tr	22.8	25.15	25.8	32.35	33.03	32.44	31.97	30.57	33.81	33.64	32.73	
Average	28.79	30.24	31.41	34.61	35.32	34.21	33.59	32.90	35.85	35.72	34.64	

Table 4.5: PSNR comparison of non-motion-compensated and motion-compensated algorithms for the corrupted input test sequences with input PSNRs of 20 dB, 25 dB and 30 dB, resp.

this improvement is reduced to approximately 0.5-0.75 dB. This can be attributed to the fact that motion causes the corresponding pixels in consecutive frames to differ which reduces the contribution of the pixel in the past frame during the filtering of the current pixel. Thus, the improvement in the output PSNR as a result of filtering is dependent upon the amount of motion and detail in the input sequence.

Table 4.5 also presents the PSNR values obtained using the motion-compensated schemes. The motion compensation has been done in the temporal direction using the half-pixel level block matching technique with a block size of 8x8. The motion vectors have been estimated using the past denoised frame and the current and the future noisy frames; the estimated motion vectors are then used for compensating motion in the past and the future noisy sequences. The proposed filter with motion compensation, MC-DCWT, yields a PSNR gain in the range 0.7-4 dB over the AWA, LMMSE and JKW filters, depending upon the amount of motion and detail in the input sequence. The more the detail in the sequence, the more significant is the effect of blurring and thus the reduction in the PSNR. The motion compensation serves to reduce the effect of motion while performing the spatio-temporal filtering and improve the PSNR gain. The applied motion compensation successfully accounts for the high amount of motion in the sequences (e.g., as in *Foreman*). With respect to the spatial denoising scheme, a PSNR gain of about 1 - 1.5 dB is achieved. The PSNR gain for the proposed filter over the spatial denoising scheme is more pronounced in this case where the motion has been compensated. Even the sequences having high motion show a significant increase in the PSNR.

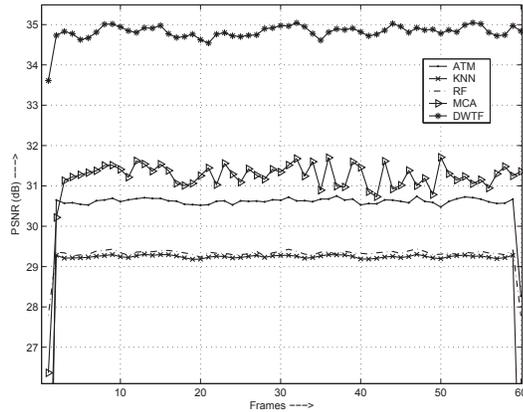
When compared to the MC3DWTr filter, the MC3DWTr filter has a marginally better performance (~ 0.25 dB) than the proposed filter, when both the filters use

a total of two frames for processing. This improvement is due to the denoised past coefficients that are used in the estimation process in the case of the MC3DWTr filter.

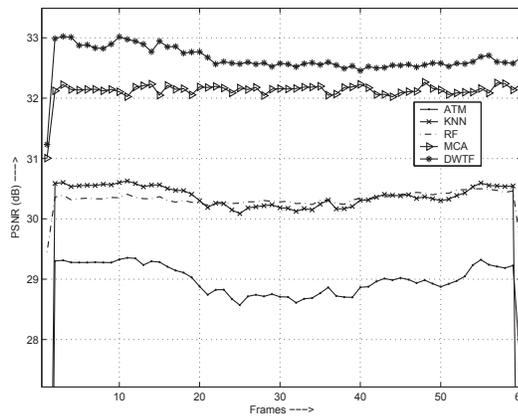
As compared to the DCWT, a gain of 0.4-0.6 dB is obtained while utilizing MC-DCWT. The PSNR gain while utilizing motion compensation is dependent upon the motion content of the input sequences. The motion compensation is able to compensate the motion in the sequences having a large amount of movement and thus, accounts for a greater increase in the output PSNR. On the other hand, the sequences which do not have a significant amount of motion do not gain much in terms of output PSNR despite the use of motion compensation. The compensation of motion in the proposed algorithm is able to suppress flicker in the filtered output and results in an output that is visually more pleasant.

The performance of the proposed filter utilizing a different number of input frames for denoising the current noisy frame is evaluated in Table 4.6. As is evident, the output PSNR of the filtered output increases with an increase in the number of input frames. The filtered outputs obtained using three frames have a PSNR improvement of about 0.5 dB over the outputs obtained using only two frames. Furthermore, the use of five frames for filtering of the current frame yields a PSNR improvement of nearly 1 dB over the use of two frames. These improvements are consistent for both the motion-compensated and non-motion-compensated filters.

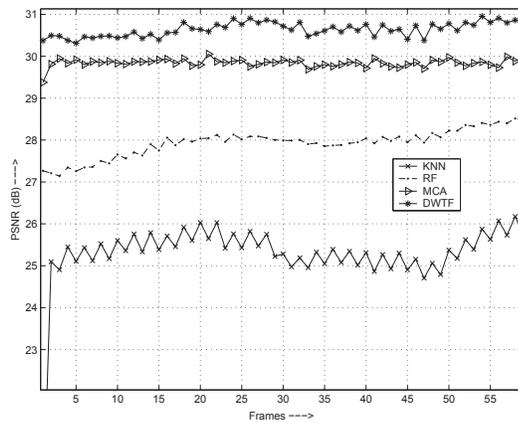
Figure 4.7 shows the PSNR curves obtained using the non-motion-compensated algorithms for 60 frames of different video sequences corrupted with different noise powers. The sequences which are used feature different amount of motion, with the *Patrol Car* sequence showing a significant amount of local as well as global motion, while the *Miss America* and *Hall* sequences showing local motion with



(a) MA: Non-motion Comp. Filtering



(b) Ha: Non-motion Comp. Filtering



(c) PC: Non-motion Comp. Filtering

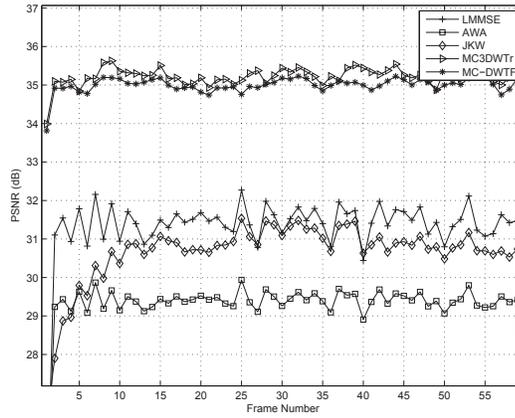
Figure 4.7: PSNR comparison of the different algorithms for 60 frames of MA, Ha and PC sequences having input PSNRs of 20 dB, 25 dB and 25 dB, resp., for non-motion compensated filtering.

	Non-motion-compensated Filters			Motion-compensated Filters			Spatial Filter
	2 Fr.	3 Fr.	5 Fr.	2 Fr.	3 Fr.	5 Fr.	(HAT)
Input PSNR = 20 dB							
MA	34.83	35.34	35.62	34.98	35.55	35.83	33.6
Ha	29.11	29.75	30.08	29.35	30.11	30.45	27.58
Fo	29.54	30.2	30.38	29.77	30.52	30.72	27.34
PC	27.29	27.63	27.92	27.85	28.07	28.35	26.84
Tr	25.88	26.19	26.42	26.43	26.77	27.12	25.31
Average	29.33	29.82	30.08	29.67	30.21	30.50	28.13
Input PSNR = 25 dB							
MA	36.78	37.25	37.52	36.93	37.54	37.86	35.82
Ha	32.47	32.87	33.37	32.71	33.36	33.70	30.99
Fo	32.15	32.67	32.93	32.44	33.10	33.31	30.75
PC	30.66	30.98	31.31	31.12	31.42	31.71	30.23
Tr	29.43	29.77	30.20	29.96	30.38	30.75	29.02
Average	32.30	32.71	33.1	32.63	33.12	33.47	31.36
Input PSNR = 30 dB							
MA	38.72	39.12	39.48	38.89	39.28	39.72	38.13
Ha	35.61	36.10	36.45	35.92	36.45	36.82	34.39
Fo	34.96	35.32	35.42	35.26	35.83	35.93	33.82
PC	34.27	34.55	34.80	34.85	35.10	35.38	34.12
Tr	33.03	33.36	33.72	33.64	33.96	34.27	32.73
Average	35.32	35.71	35.97	35.72	36.13	36.43	34.64

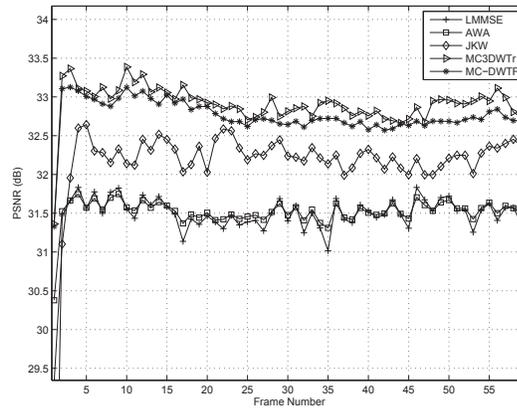
Table 4.6: PSNR comparison of the proposed non-motion-compensated and motion-compensated algorithms with 2, 3 and 5 corrupted input frames with input PSNRs of 20dB, 25 dB and 30 dB, resp.

a stationary background. The performance of the KNN and RF filters suffers significantly while processing frames containing significant amount of motion and even in the presence of low motion, their performance is not aggressive enough to remove sufficient amount of noise from the video. The proposed filter outperforms these filters to the tune of 4-6 dB. Even compared to the wavelet domain filter (MCA), the proposed filter yields an improvement of about 0.5-3 dB.

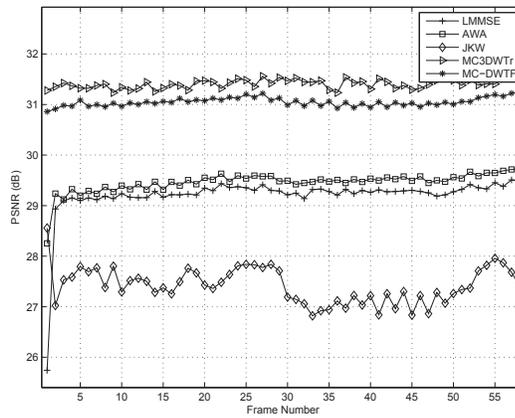
The PSNR curves obtained using motion-compensated filtering on the above sequence are presented in Figure 4.9. As compared to the AWA, LMMSE and



(a) MA: Motion-comp. Filtering

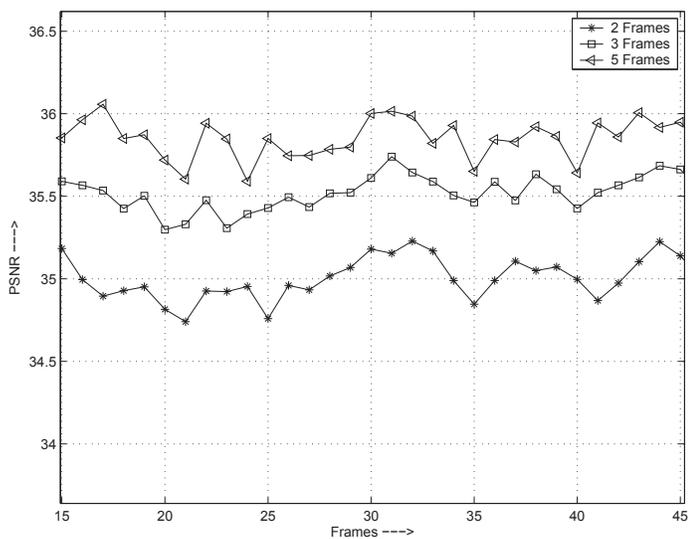


(b) Ha: Motion Comp. Filtering

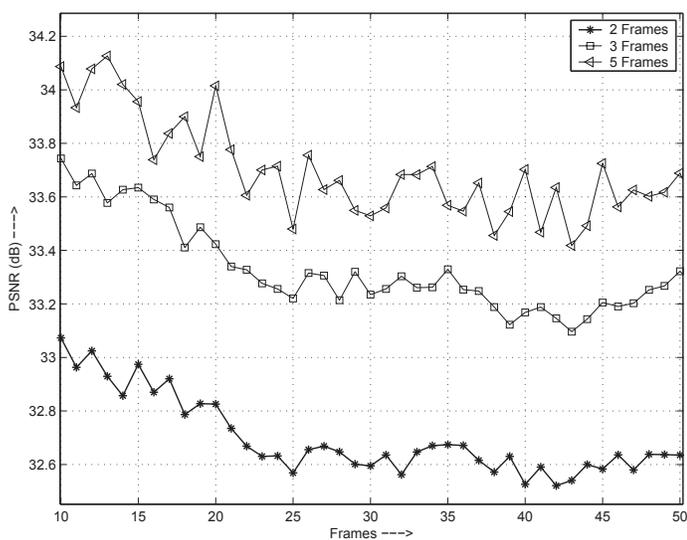


(c) PC: Motion Comp. Filtering

Figure 4.8: PSNR comparison of the different algorithms for 60 frames of MA, Ha and PC sequences having input PSNRs of 20 dB, 25 dB and 25 dB, resp., for Motion compensated filtering



(a) MA: MC-DCWT using different number of frames



(b) Ha: MC-DCWT using different number of frames

Figure 4.9: PSNR comparison for MA and Ha sequences having input PSNRs of 20 dB and 25 dB, resp., using MC-DCWT for 2, 3 and 5 input frames

JKW schemes, the proposed filter yields an improvement of about 1.5 -2 dB. The performance of the MC3DWTr filter is ~ 0.25 dB better than the proposed filter on average.

The PSNR curves comparing the proposed filter with different number of input frames are presented in the Fig. 4.9. The PSNR improvement achieved with the use of additional frames for filtering the current frame is evident from these curves. The increased number of frames for filtering provides additional data for processing purposes and this in turn leads to more aggressively filtered outputs.

Figure 4.10 presents the performance results on the 17th frame of the *Miss America* sequence which has an input PSNR of 20 dB using filters without motion-compensation. The proposed DCWT filter yields the best output in terms of visual quality as compared to ATM, KNN, RF and MCA filters. The time domain filters cause excessive blurring and are also unable to filter out the noise aggressively. Even the wavelet domain filter, MCA, causes the loss of some of the detail in the denoised image and is unable to reduce noise to a good extent. The proposed filter, on the other hand, is able to preserve all the fine spatial details and also reduce the amount of flicker and residual noise in the actual video. The visual performance for the motion-compensated schemes is compared in Fig. 4.11. The superior performance of the proposed MC-DCWT filter in terms of edge preservation and noise reduction is clearly evident as compared to the AWA , LMMSE and JKW filters. These time domain filters used for comparison are unable to suppress the noise to the same extent as the proposed filter. The proposed filter performs yields almost similar quality results when compared visually with the MC3DWTr filter.

Figure 4.12 presents the filtered outputs for the 5th frame of the *Hall* video sequence. The increased number of frames for filtering provides additional data for processing purposes and this in turn leads to more aggressively filtered outputs.

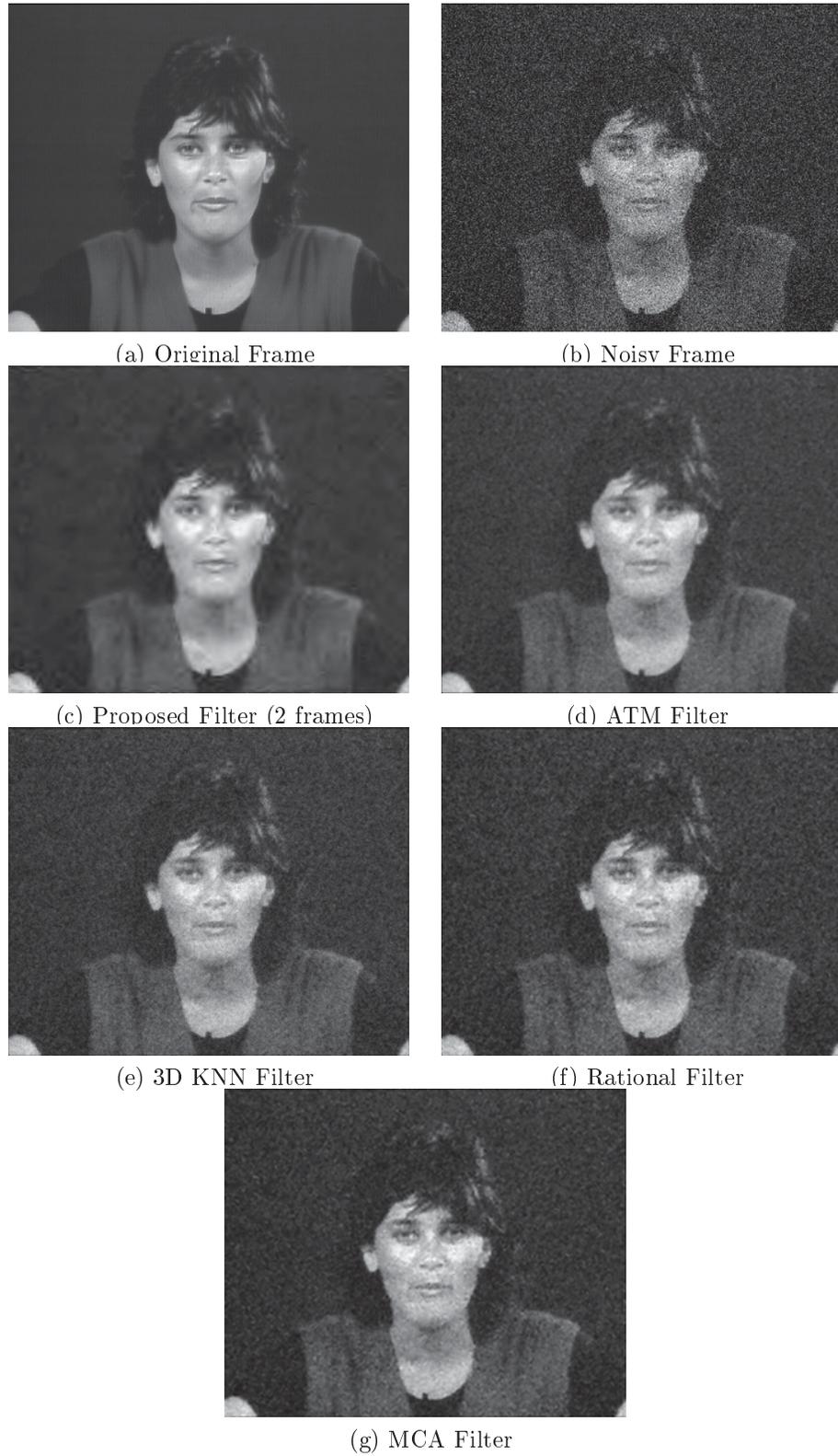


Figure 4.10: Filtered outputs of the 17th frame of the *Miss America* sequence for non-motion-compensated filters

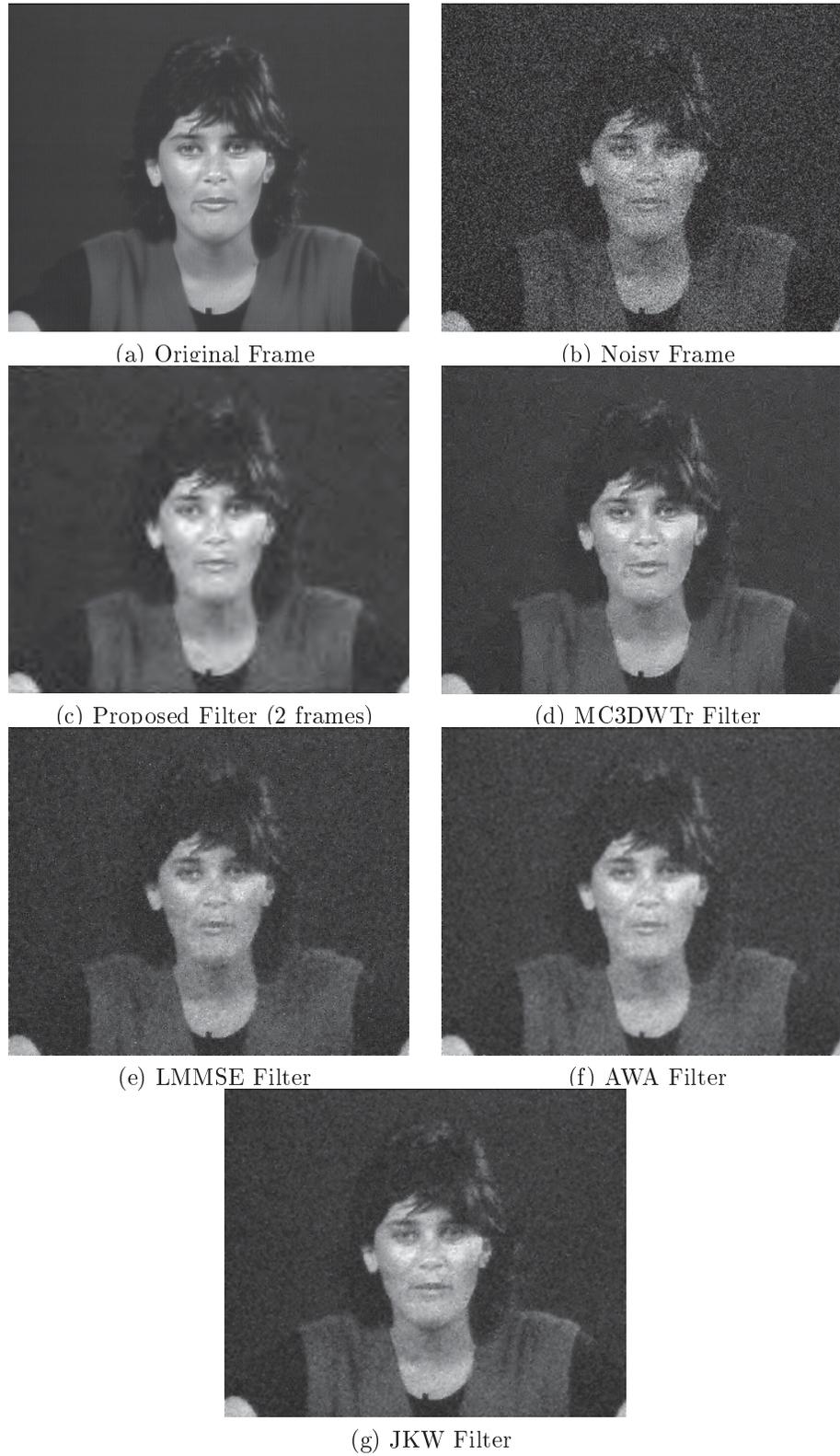


Figure 4.11: Filtered outputs of the 17th frame of the *Miss America* sequence for motion-compensated filters

The sequences produced by the filter using five input frames are more visually pleasing and have a reduced amount of flicker artifacts. Furthermore, the edges are better preserved while at the same time, noise is more aggressively reduced.

4.7 Conclusion

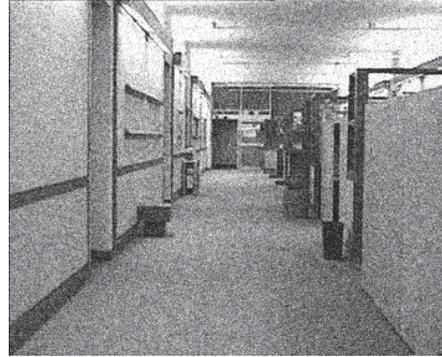
The problem of video noise reduction in the wavelet domain has been addressed in this chapter. The spatial as well as temporal redundancies have been exploited for the development of a low-complexity and adaptive technique which operates under the video modeling framework proposed in this dissertation. Before the subband coefficients are processed spatially for noise reduction, their correlation with the corresponding temporal neighbors has been minimized by the use of DCT in the temporal direction. It has been shown that the decorrelated coefficients can still be modeled using the GG distribution and the spatial inter-subband correlations which existed before the application of the DCT, still persist after the decorrelation. Using this, the spatial noise reduction has been performed using a locally adaptive, hierarchical noise reduction technique.

The effect of explicit motion estimation as well as compensation on the performance of the proposed technique has been evaluated and discussed. Furthermore, the effect of using a variable support in the temporal direction (2, 3 or 5 frames) while employing the proposed technique has also been analyzed.

It has also been shown using extensive experimentation that the proposed scheme outperforms several state-of-the-art video noise reduction algorithms, in time as well as wavelet domains, quantitatively as well as in terms of visual quality.



(a) Original Frame



(b) Noisy Frame



(c) Proposed Filter (2 Frames)



(d) Proposed Filter (3 Frames)



(e) Proposed Filter (5 Frames)

Figure 4.12: Filtered outputs of the 5th field of the Hall sequence for proposed MC-DCWT filter with 2, 3 and 5 frames, respectively.

Chapter 5

Bayesian Algorithms for Video Noise Reduction in the Wavelet Domain

5.1 Introduction

In the previous chapter, the problem of additive white Gaussian noise removal from the video sequences was considered. In order to aggressively reduce noise while still preserving the details in the video frames, a spatio-temporal filter was formulated within the proposed wavelet-domain video modeling framework (see Chapter 3) where the temporal redundancies between consecutive frames in the wavelet domain were minimized before spatially denoising the subband coefficients. While the approach used in the previous chapter comprised of minimizing the temporal correlations, it is also possible to utilize those correlations and take those into account while denoising the current subband coefficients.

Thus, in this chapter, Bayesian operators are developed which yield an esti-

mate of the noise-free wavelet coefficients in the current frame conditioned not only on the available noisy coefficients belonging to the current frame, but also on the corresponding estimated noise-free coefficients in the past frame. A spatio-temporal framework for the modeling of the subband coefficients was discussed in Chapter 3 that not only models the current subband coefficients using the generalized Gaussian distribution, but also the difference between the neighboring subbands, which is caused due to motion, using the same *a priori* distribution and this framework is exploited for this development. Minimum mean squared error (MMSE) estimation as well as maximum *a posteriori* (MAP) estimation is considered for the development of the filtering operators in the Bayesian framework.

The chapter is organized as follows. A brief formulation of the problem is provided in Section 5.2. The Bayesian operator based on MMSE estimation is discussed in Section 5.3, while the one based on MAP estimation is developed in Section 5.4. Section 5.5 presents the simulation results and a detailed performance comparison of the proposed algorithm with some of the existing state-of-the-art techniques for video denoising.

5.2 Problem Formulation

Consider a noise-free video sequence whose n^{th} frame can be represented as \mathbf{f}^n . Each of the noise-free frames in the original video sequence is corrupted with a set of zero-mean AWGN, η^n , having a variance σ_η^2 , so as to yield the observed frames of the noisy video sequence:

$$\mathbf{g}^n = \mathbf{f}^n + \eta^n \quad (5.1)$$

This chapter is concerned with obtaining an estimate, $\hat{\mathbf{f}}^n$, of the noise-free frame, \mathbf{f}^n , given the observed noisy frame, \mathbf{g}^n .

The input noisy frame is transformed to the wavelet domain using the critically-sampled discrete wavelet transform (DWT) and owing to the linearity and orthogonality of the wavelet transform, the corrupting noise in the wavelet domain is also additive:

$$\mathbf{Y}^n = \mathbf{X}^n + \epsilon^n \quad (5.2)$$

where \mathbf{Y}^n and \mathbf{X}^n are the sets of noisy and noise-free wavelet coefficients and ϵ^n is the set of corrupting zero-mean AWGN coefficients having a variance σ_ϵ^2 . Owing to the above mentioned orthogonality and linearity of the wavelet transform, $\sigma_\epsilon^2 = \sigma_\eta^2$.

Furthermore, as discussed in Section 3.3, the noise-free wavelet coefficients in the current frame n , \mathbf{X}^n , can be expressed in terms of the noise-free coefficients in the previous frame $n - 1$, \mathbf{X}^{n-1} , as

$$\mathbf{X}^n = \mathbf{X}^{n-1} + \mathbf{M}^{n-1} \quad (5.3)$$

where \mathbf{M}^{n-1} is the motion noise introduced in the subband representation of the frame at instance $n - 1$.

Thus, in this chapter we address the wavelet domain noise reduction problem of finding an estimate $\hat{\mathbf{X}}^n$ of the original frame \mathbf{X}^n based on the noisy observation $\hat{\mathbf{Y}}^n$ where \mathbf{X}^n has been modeled in terms of the previous frame \mathbf{X}^{n-1} and the motion noise .

5.3 Spatio-temporal MMSE Estimation

To estimate the noise-free wavelet coefficients corresponding to frame n , a Bayesian processor based on the signal models described in the previous section is used for each subband. Since the past frame $n - 1$ is denoised ahead of the frame n , the noise-free estimates of the wavelet coefficients in that frame are available for estimation process, along with the noisy coefficients in the current frame. For the representation of the Bayesian processor in each subband, we denote the noisy coefficients in the subband as y^n , the noise-free estimates of the wavelet coefficients in the current and corresponding subband in the past frame as \hat{x}^n and \hat{x}^{n-1} , respectively. The difference set for a subband in frame n is denoted by m^{n-1} .

The Bayes risk estimator under a quadratic cost function minimizes the mean-square error (MSE) and is given by the conditional mean of x^n , given y^n and \hat{x}^{n-1} , and can be expressed as:

$$\hat{x}^n(y^n, \hat{x}^{n-1}) = \int x^n p_{x^n|y^n, \hat{x}^{n-1}}(x^n|y^n, \hat{x}^{n-1}).dx^n \quad (5.4)$$

Resolving (5.4), it can be expressed as

$$\hat{x}^n(y^n, \hat{x}^{n-1}) = \frac{\int x^n p_{y^n|x^n}(y^n|x^n)p_{x^n|\hat{x}^{n-1}}(x^n|\hat{x}^{n-1}).dx^n}{\int p_{y^n|x^n}(y^n|x^n)p_{x^n|\hat{x}^{n-1}}(x^n|\hat{x}^{n-1}).dx^n} \quad (5.5)$$

Note that in (5.5), the noisy coefficients in the current frame have been modeled as the sum of the noise-free coefficients in the current frame and AWGN (i.e., $y^n = x^n + \epsilon^n$), and thus

$$p_{y^n|x^n}(y^n|x^n) = p_{\epsilon^n}(y^n - x^n). \quad (5.6)$$

Here, ϵ^n is the random variable denoting the AWGN having a zero mean and variance, σ_ϵ^2 , which can be estimated as [45]

$$\hat{\sigma}_\epsilon^2 = \left(\frac{\text{median}(|HH_1|)}{0.6745} \right)^2 \quad (5.7)$$

where HH_1 denotes the largest diagonal detail subband in the wavelet decomposition of the noisy frame.

Also, using the conditional description of the current frame subband coefficients in terms of the previous frame and motion noise coefficients as described in (5.3) (i.e., $x^n = \hat{x}^{n-1} + m^{n-1}$),

$$p_{x^n|\hat{x}^{n-1}}(x^n|\hat{x}^{n-1}) = p_{m^{n-1}}(x^n - \hat{x}^{n-1}). \quad (5.8)$$

where the set of random variables, m^{n-1} has been shown to have a GG distribution (see Section 3.3) and can be described as:

$$p_m(m) = \frac{\nu_m}{2s_m\Gamma(\frac{1}{\nu_m})} \exp(-|\frac{x}{s_m}|^{\nu_m}), \quad s_m, \nu_m > 0 \quad (5.9)$$

To estimate the model parameters ν_m and s_m for the set of random variables m^{n-1} , we obtain a noisy version of this random variable by subtracting the estimated wavelet coefficients in the past frame, \hat{x}^{n-1} , from the observed wavelet coefficients in the current frame, y^n . This results in a noisy mixture comprising of two independent random variables, one GG distributed (m^{n-1}) and one Gaussian distributed (ϵ^n). The model parameters ν_m and s_m can thus be estimated from this noisy mixture using the following equations [48]:

$$\sigma_w^2 = \sigma_\epsilon^2 + \frac{s_m^2\Gamma(\frac{3}{\nu_m})}{\Gamma(\frac{1}{\nu_m})}, \quad (5.10)$$

$$m_{4,w} = 3\sigma_\epsilon^4 + \frac{6\sigma_\epsilon^2 s_m^2 \Gamma(\frac{3}{\nu_m})}{\Gamma(\frac{1}{\nu_m})} + \frac{s_m^4 \Gamma(\frac{5}{\nu_m})}{\Gamma(\frac{1}{\nu_m})} \quad (5.11)$$

where σ_w^2 and $m_{4,w}$ are the variance and the fourth moments of the noisy version of the random variable, m^{n-1} , respectively.

The resultant Bayesian processor for each subband can thus be written as

$$\hat{x}^n(y^n, \hat{x}^{n-1}) = \frac{\int x^n p_{\epsilon^n}(y^n - x^n) p_{m^{n-1}}(x^n - \hat{x}^{n-1}) . dx^n}{\int p_{\epsilon^n}(y^n - x^n) p_{m^{n-1}}(x^n - \hat{x}^{n-1}) . dx^n}. \quad (5.12)$$

For the first frame, where the past frame coefficients are not available, the estimation is performed using only the available noisy coefficients. In this case, the Bayesian operator can be expressed as:

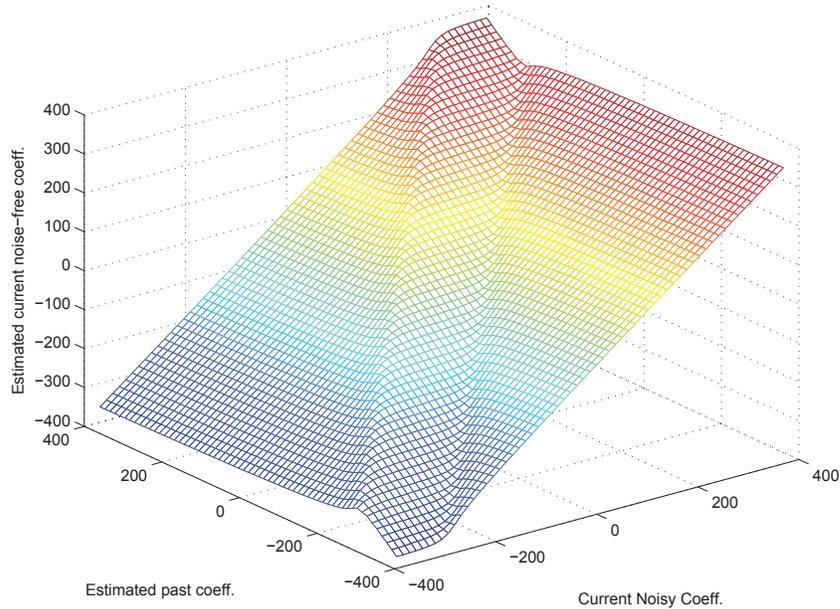
$$\hat{x}^n(y^n) = \int x^n p_{x^n|y^n}(x^n|y^n) . dx^n \quad (5.13)$$

which, using (5.6) can be rewritten as

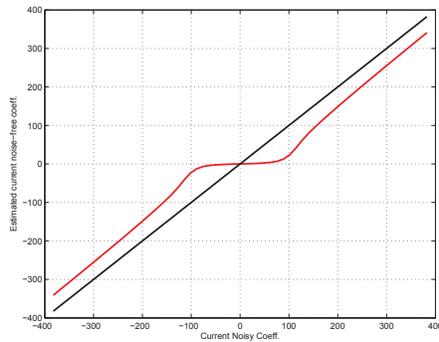
$$\hat{x}^n(y^n) = \frac{\int x^n p_{\epsilon^n}(y^n - x^n) p_{x^n}(x^n) . dx^n}{\int p_{\epsilon^n}(y^n - x^n) p_{x^n}(x^n) . dx^n} \quad (5.14)$$

where $p_{x^n}(x^n)$ has a GG distribution (see Section 3.2) and whose model parameters can be estimated using (5.10) and (5.11) when the available noisy coefficients in the subband, y^n , are used to estimate the variance and the fourth moment.

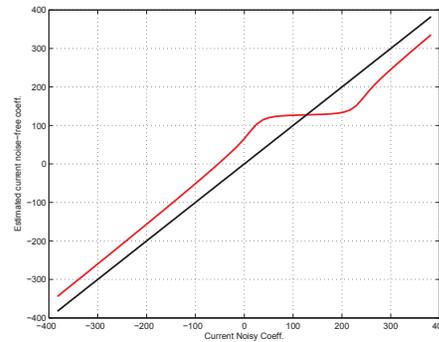
Figures 5.1 (a) and 5.2 (a) show the resultant shrinkage function, where the Bayesian output is conditioned upon two input parameters. The shrinkage function considers both the noisy current frame coefficient and the estimated noise-free past frame coefficient to estimate the corresponding noise-free coefficient in the current frame. If the estimated value of the coefficient in the past frame is high (or low), then owing to the redundancy in time, the value of the noise-free co-



(a) Shrinkage Function ($\sigma_\epsilon = 20$, $\nu_m = 0.75$, $s_m = 2$)

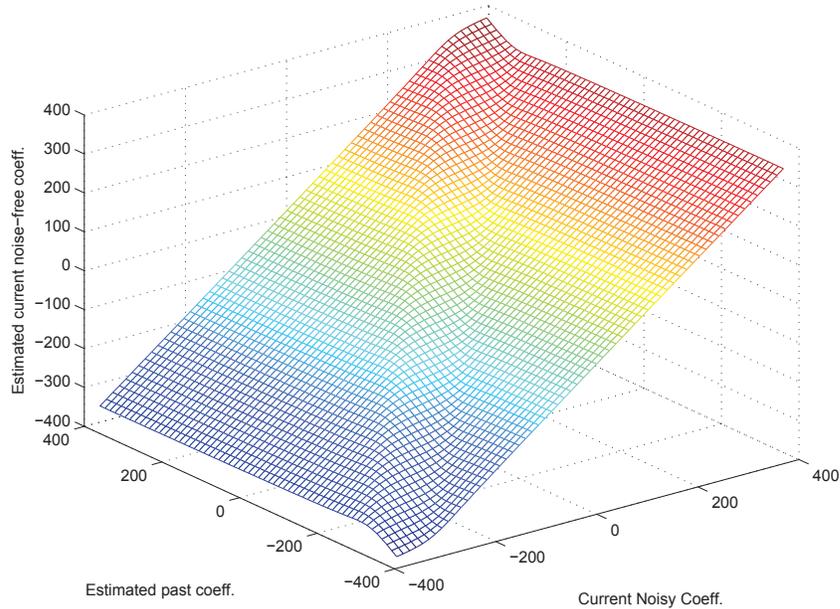


(b) Shrinkage function for constant $\hat{x}^{n-1}(= 0)$

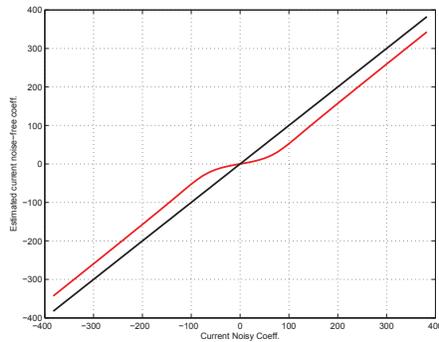


(c) Shrinkage function for constant $\hat{x}^{n-1}(> 0)$

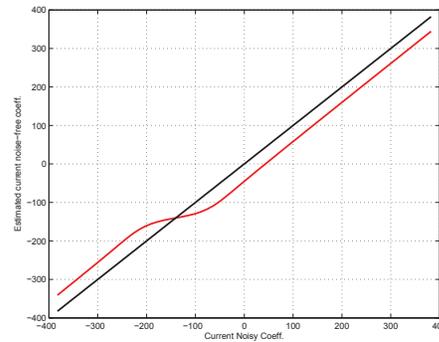
Figure 5.1: Shrinkage function obtained using the MMSE estimation of the current noise-free coefficients conditioned upon the current noisy coefficients and the denoised past frame coefficients. Also shown are the shrinkage operation as functions of the noisy subband coefficients in the current frame for a particular value of \hat{x}^{n-1} .



(a) Shrinkage Function ($\sigma_\epsilon = 25$, $\nu_m = 0.9$, $s_m = 10$)



(b) Shrinkage function for constant $\hat{x}^{n-1}(= 0)$



(c) Shrinkage function for constant $\hat{x}^{n-1}(< 0)$

Figure 5.2: Shrinkage function obtained using the MMSE estimation of the current noise-free coefficients conditioned upon the current noisy coefficients and the denoised past frame coefficients. Also shown are the shrinkage operation as functions of the noisy subband coefficients in the current frame for a particular value of \hat{x}^{n-1} .

efficient in the current frame has a high probability of being a large (or small) value. This intuitive reasoning is captured in the presented shrinkage function. This can be further illustrated by considering particular values of the estimated noise-free coefficient in the past frame, \hat{x}^{n-1} , for which the operator in (5.12) reduces to the Bayesian function which is conditioned on only one parameter, y^n , and is similar to (5.14). This one variable function is depicted in Figures 5.1 (b-c) and 5.2 (b-c) where the shrinkage operator has been plotted only as a function of y^n for a constant value of \hat{x}^{n-1} . For $\hat{x}^{n-1} = 0$, the shrinkage operators are similar to the traditional MMSE operators. However, when $\hat{x}^{n-1} > 0$ or $\hat{x}^{n-1} < 0$, then the shrinkage of y^n is not symmetric about 0, but is conditioned upon the value of \hat{x}^{n-1} . This allows the proposed MMSE shrinkage operator to be adaptive to not only the input current frame noisy data, but also to the estimated noise-free values from the past frames.

The output of the inverse wavelet transform operation on the denoised coefficients in the current frame yields the denoised time domain pixel values as

$$\hat{\mathbf{f}}^n = \text{InverseDWT}(\hat{\mathbf{X}}^n). \quad (5.15)$$

5.4 Spatio-temporal MAP Estimation

In this section, the Bayesian processor based on the MAP estimation is developed for the estimation of the noise-free wavelet coefficients corresponding to frame n . As in the previous section, for the representation of the Bayesian processor in each subband, we denote the noisy coefficients in the subband as y^n , the noise-free estimates of the wavelet coefficients in the current and corresponding subband in the past frame as x^n and \hat{x}^{n-1} , respectively. The difference set for a subband in frame n is denoted by m^{n-1} . The noise-free estimates of the wavelet coefficients

in the past frame $n - 1$ are used along with the noisy coefficients in the current frame in the estimation process.

The MAP estimate of the noise-free subband coefficient can be expressed as

$$\frac{\partial}{\partial x^n} \ln p_{x^n|y^n, \hat{x}^{n-1}}(x^n|y^n, \hat{x}^{n-1}) = 0, \quad (5.16)$$

where the density of the current frame noise-free subband coefficients, x^n , conditioned on y^n and \hat{x}^{n-1} can be expressed as

$$p_{x^n|y^n, \hat{x}^{n-1}}(x^n|y^n, \hat{x}^{n-1}) = \frac{p_{y^n|x^n}(y^n|x^n)p_{x^n|\hat{x}^{n-1}}(x^n|\hat{x}^{n-1})}{\int p_{y^n|x^n}(y^n|x^n)p_{x^n|\hat{x}^{n-1}}(x^n|\hat{x}^{n-1}).dx^n}. \quad (5.17)$$

Using (5.6) and (5.8), (5.18) can be rewritten as

$$p_{x^n|y^n, \hat{x}^{n-1}}(x^n|y^n, \hat{x}^{n-1}) = \frac{p_{\epsilon^n}(y^n - x^n)p_{m^{n-1}}(x^n - \hat{x}^{n-1})}{\int p_{\epsilon^n}(y^n - x^n)p_{m^{n-1}}(x^n - \hat{x}^{n-1}).dx^n}. \quad (5.18)$$

Since ϵ^n is the random variable denoting the AWGN having a mean zero and variance σ_ϵ^2 , which can be estimated using (5.7), and m^{n-1} has a GG distribution with parameters ν_m and s_m , substituting (5.18) in (5.16),

$$\frac{\partial}{\partial x^n} \ln \left[\frac{p_{\epsilon^n}(y^n - x^n)p_{m^{n-1}}(x^n - \hat{x}^{n-1})}{\int p_{\epsilon^n}(y^n - x^n)p_{m^{n-1}}(x^n - \hat{x}^{n-1}).dx^n} \right] = 0. \quad (5.19)$$

Solving (5.19),

$$\frac{\partial}{\partial x^n} \ln [p_{\epsilon^n}(y^n - x^n)] + \frac{\partial}{\partial x^n} \ln [p_{m^{n-1}}(x^n - \hat{x}^{n-1})] = 0, \quad (5.20)$$

which results in:

$$\left(\frac{y^n - x^n}{\sigma_\epsilon^2} \right) - \frac{\partial}{\partial x^n} \left| \frac{x^n - \hat{x}^{n-1}}{s_m} \right|^{\nu_m} = 0. \quad (5.21)$$

The solution of (5.21) in terms of x^n is the MAP estimate, \hat{x}^n , of the noise-free subband coefficient. For the first frame, where the past frame coefficients are not available, the estimation is performed using only the available noisy coefficients. In this case, the Bayesian MAP operator can be expressed in its traditional form as:

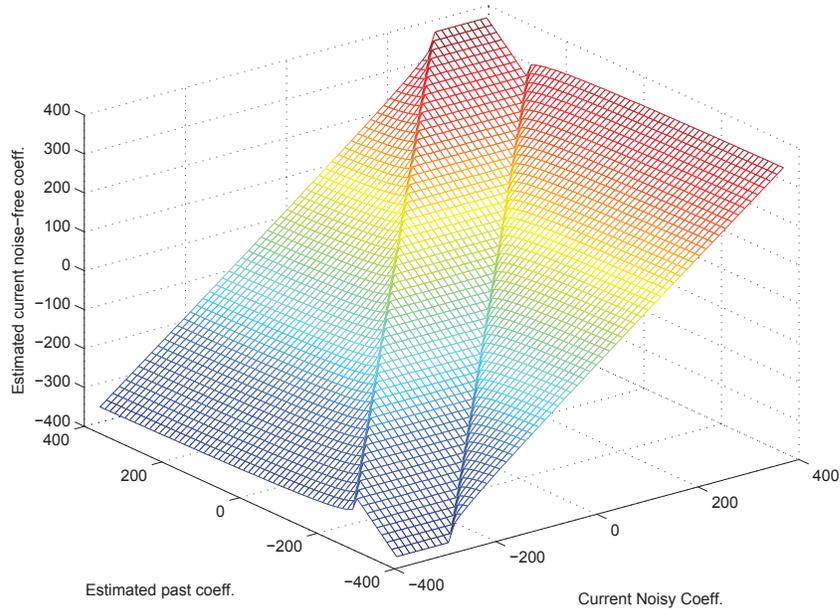
$$\frac{\partial}{\partial x^n} \ln p_{x^n|y^n}(x^n|y^n) = 0, \quad (5.22)$$

which can be solved as:

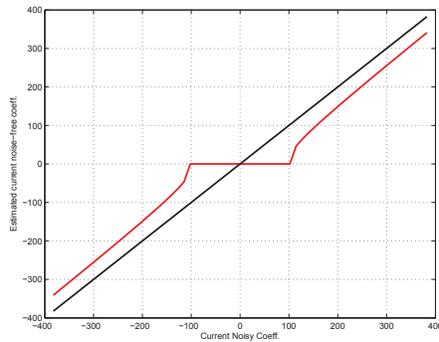
$$\left(\frac{y^n - x^n}{\sigma_\epsilon^2} \right) - \frac{\partial}{\partial x^n} \left| \frac{x^n}{s_x} \right|^{\nu_x} = 0 \quad (5.23)$$

where ν_x and s_x are the model parameters for $p_{x^n}(x^n)$, which has a GG density, and can be estimated using (5.10) and (5.11) when the available noisy coefficients in the subband are used to estimate σ_w^2 and $m_{4,w}$.

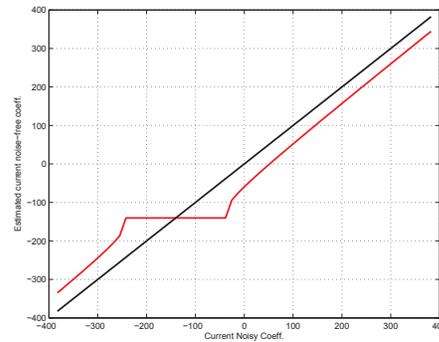
Figures 5.3 (a) and 5.4 (a) show the proposed shrinkage operator for the MAP Bayesian estimation conditioned upon two input parameters. The shrinkage function considers both y^n and the estimated noise-free past frame coefficients, \hat{x}^{n-1} , in order to estimate the corresponding noise-free coefficient in the current frame, \hat{x}^n . Owing to the high redundancy between the corresponding coefficients in successive frames, the probability of a high (or low) valued coefficient in the current frame is dependent upon the value of the corresponding coefficient in the past frame. The presented 2D shrinkage operators illustrate this reasoning. This can be further simplified by considering particular values of the estimated noise-free coefficient in the past frame, \hat{x}^{n-1} , for which the operator in (5.21) reduces to the traditional conditional Bayesian function dependent only upon y^n and is similar to (5.23). This one variable function is depicted in Figures 5.3 (b-c) and 5.4 (b-c) where the shrinkage operator has been plotted only as a function of y^n for a



(a) MAP Shrinkage Function ($\sigma_\epsilon = 20$, $\nu_m = 0.75$, $s_m = 2$)

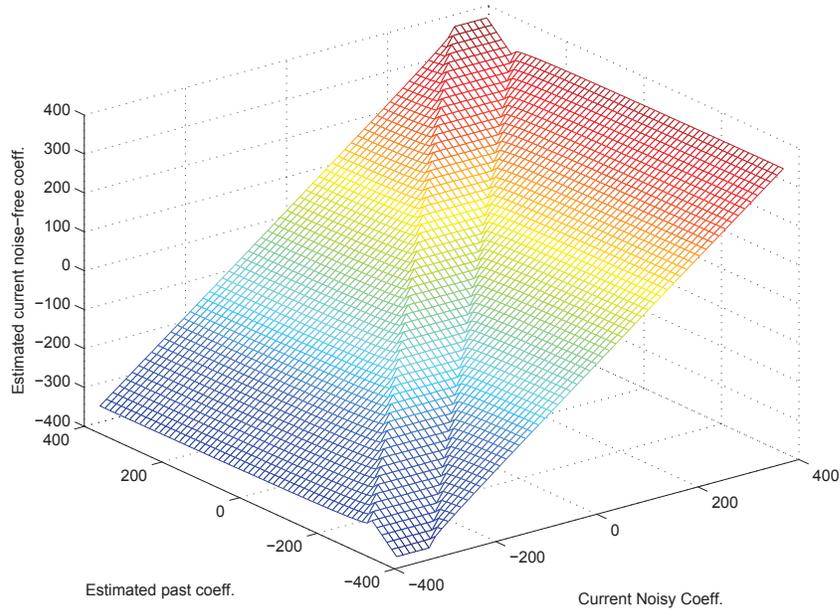


(b) Shrinkage function for constant $\hat{x}^{n-1}(= 0)$

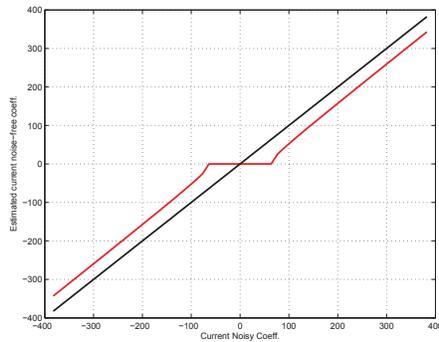


(c) Shrinkage function for constant $\hat{x}^{n-1}(< 0)$

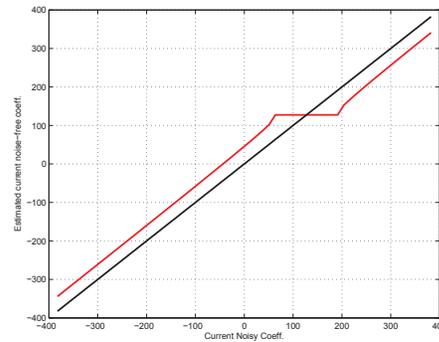
Figure 5.3: Shrinkage function obtained using the MAP estimation of the current noise-free coefficients conditioned upon the current noisy coefficients and the denoised past frame coefficients. Also shown are the shrinkage operation as functions of the noisy subband coefficients in the current frame for a particular value of \hat{x}^{n-1} .



(a) MAP Shrinkage Function ($\sigma_\epsilon = 25$, $\nu_m = 0.9$, $s_m = 10$)



(b) Shrinkage function for constant $\hat{x}^{n-1}(= 0)$



(c) Shrinkage function for constant $\hat{x}^{n-1}(> 0)$

Figure 5.4: Shrinkage function obtained using the MAP estimation of the current noise-free coefficients conditioned upon the current noisy coefficients and the denoised past frame coefficients. Also shown are the shrinkage operation as functions of the noisy subband coefficients in the current frame for a particular value of \hat{x}^{n-1} .

constant value of \hat{x}^{n-1} . For $\hat{x}^{n-1} = 0$, the traditional MAP operators which are symmetric about 0 along the x- and the y- axis are obtained. However, when $\hat{x}^{n-1} > 0$ or $\hat{x}^{n-1} < 0$, then the shrinkage of y^n is not symmetric about 0, but is conditioned upon the value of \hat{x}^{n-1} and thus, the point of symmetry slides along the $x = y$ line based on \hat{x}^{n-1} . This allows for a spatio-temporally adaptive MAP operator which can reduce noise in the subband coefficients more intuitively.

The output of the inverse wavelet transform operation on the denoised coefficients in the current frame yields the denoised time domain pixel values, $\hat{\mathbf{f}}^n = \text{InverseDWT}(\hat{\mathbf{X}}^n)$.

5.5 Experimental Results and Discussion

The performance of the proposed algorithms, TMMSE and TMAP, has been evaluated using the following state-of-the-art algorithms:

- in the time-domain: 3D Rational filter (RF) [26], 3D KNN filter [24], 3D α -trimmed mean (ATM) filter [24], 3D LMMSE filter [35], the adapted weighted averaging (AWA) filter [35] and the joint Kalman and Wiener filter (JKW) [39].
- in the wavelet-domain: multiple class averaging (MCA) algorithm [29] and MC3DWTr [58].

As described in Section 4.6.1, peak signal-to-noise ratio (PSNR) and visual quality are used as the performance criteria for the relative evaluation of the different algorithms.

Five different test sequences, viz. *Hall* (Ha), *Miss America* (MA), *Foreman* (Fo), *Patrol Car* (PC) and *Train* (Tr), corrupted by three different AWGN powers have been used for evaluation. The input PSNRs of the corrupted sequences are

20 dB, 25 dB and 30 dB. The results presented for the proposed algorithms are obtained after compensating the motion in the input sequence. The motion compensation is done in the wavelet domain using the half-pixel level block matching technique with a block size of 8×8 . The motion vectors are estimated using the past denoised frame and the current noisy frame; the estimated motion vectors are then used for compensating motion in the past and the future noisy sequences. The Symmlet-8 wavelet is used to decompose the input frames upto 4 dyadic levels.

Table 5.1 presents the average PSNR values for the different algorithmic outputs with the sequences having PSNRs of 20 dB, 25 dB and 30 dB as input. The proposed filters, TMMSE and TMAP, yield an average PSNR gain in the range of 6 to 14 dB over the corrupted sequences and gains of 1-7 dB over the compared schemes. The amount of increase observed in the PSNR is dependent upon the amount of detail in the sequence as well as the amount of motion occurring in the captured scene. The more the amount of fine detail in the scene, the more likely it is that some of the detail is smoothed over as a result of the filtering, resulting in a reduced output PSNR. The effect of motion can be explained by the fact that motion causes the corresponding pixels in consecutive frames to differ, which reduces the contribution of the pixel in the past frame during the filtering of the current pixel. The motion compensation that is applied serves to reduce the effect of motion while performing the spatio-temporal filtering and improves the PSNR gain. It successfully accounts for the high amount of motion in the sequences (e.g., as in Fo, PC and Tr). The proposed schemes even outperform the corresponding wavelet domain algorithm, MCA, to the tune of 2-3 dB. When compared to the state-of-the-art MC3DWTr, the proposed techniques, TMMSE and TMAP, yield an average improvement of ~ 0.75 dB and ~ 0.5 dB, respectively. Only for the MA

	ATM	KNN	RF	MCA	LMMSE	AWA	JKW	MC3DWT _{Tr}	TMMSE	TMAP
Input PSNR = 20 dB										
MA	30.37	29.0	29.29	31.21	30.96	29.01	30.65	35.35	34.10	33.62
Ha	27.66	27.29	27.31	28.09	27.35	27.75	28.68	29.57	31.19	30.76
Fo	27.96	27.75	28.11	28.07	26.14	28.09	27.64	29.85	30.94	30.64
PC	22.94	24.14	25.97	26.15	25.5	25.6	24.89	27.99	28.95	28.63
Tr	22.38	23.89	24.27	25.53	25.3	25.22	25.59	26.72	27.60	27.38
Average	26.25	26.41	26.95	27.81	27.05	27.15	27.48	29.89	30.55	30.21
Input PSNR = 25 dB										
MA	33.61	33.45	34.18	34.92	34.57	33.36	32.22	37.32	36.35	36.04
Ha	28.74	29.98	30.34	32.04	30.86	31.07	32.37	32.92	34.2	33.89
Fo	30.58	30.37	31.66	31.44	30.13	29.85	29.42	32.59	33.65	33.37
PC	23.06	25.18	27.93	29.95	28.83	29.08	27.41	31.31	32.23	31.95
Tr	22.9	24.8	25.43	29.25	28.74	28.6	28.25	30.08	31.15	30.92
Average	27.78	28.75	29.91	31.52	30.63	30.39	29.97	32.84	33.51	33.23
Input PSNR = 30 dB										
MA	37.69	37.19	37.53	37.78	37.66	37.3	37.25	38.97	38.75	38.4
Ha	29.2	31.38	31.64	35.13	34.66	34.97	35.11	36.05	37.3	37.05
Fo	31.25	32.00	33.44	34.65	33.95	31.15	31.72	35.41	36.42	36.18
PC	23.06	25.48	28.62	33.2	32.38	32.55	29.85	34.98	36.08	35.9
Tr	22.8	25.15	25.8	32.35	32.44	31.97	30.57	33.81	34.75	34.55
Average	28.79	30.24	31.41	34.61	34.21	33.59	32.90	35.85	36.66	36.41

Table 5.1: PSNR comparison of the proposed algorithms, TMMSE and TMAP, with several others for the corrupted input test sequences with input PSNRs of 20 dB, 25 dB and 30 dB, resp.

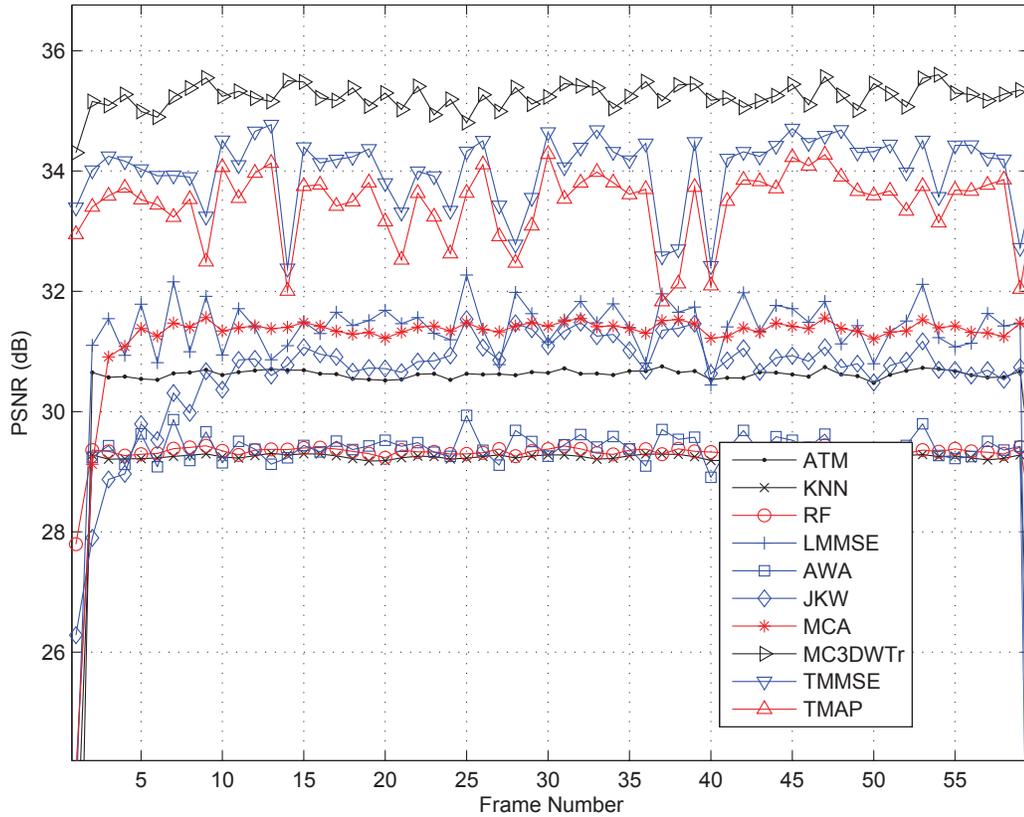


Figure 5.5: PSNR comparison of the different algorithms for *Miss America* sequence having an input PSNR of 20 dB

sequence does the MC3DWTr outperform the proposed techniques. This can be attributed to the very low amount of movement in the sequence along with a stationary background and this allows for the explicit application of the DWT in the temporal direction to gain an advantage over spatio-temporal modeling.

Among the two schemes proposed in this chapter, TMMSE outperforms TMAP by $\sim 0.2-0.3$ dB on average. This can be attributed to the smooth shrinkage functions generated by the LMMSE operator as compared to the soft thresholding which is performed by the operator developed by the MAP mechanism.

Figures 5.5, 5.6 and 5.7 show the PSNR curves obtained using the different algorithm when processing MA, Ha and PC sequences corrupted to have input

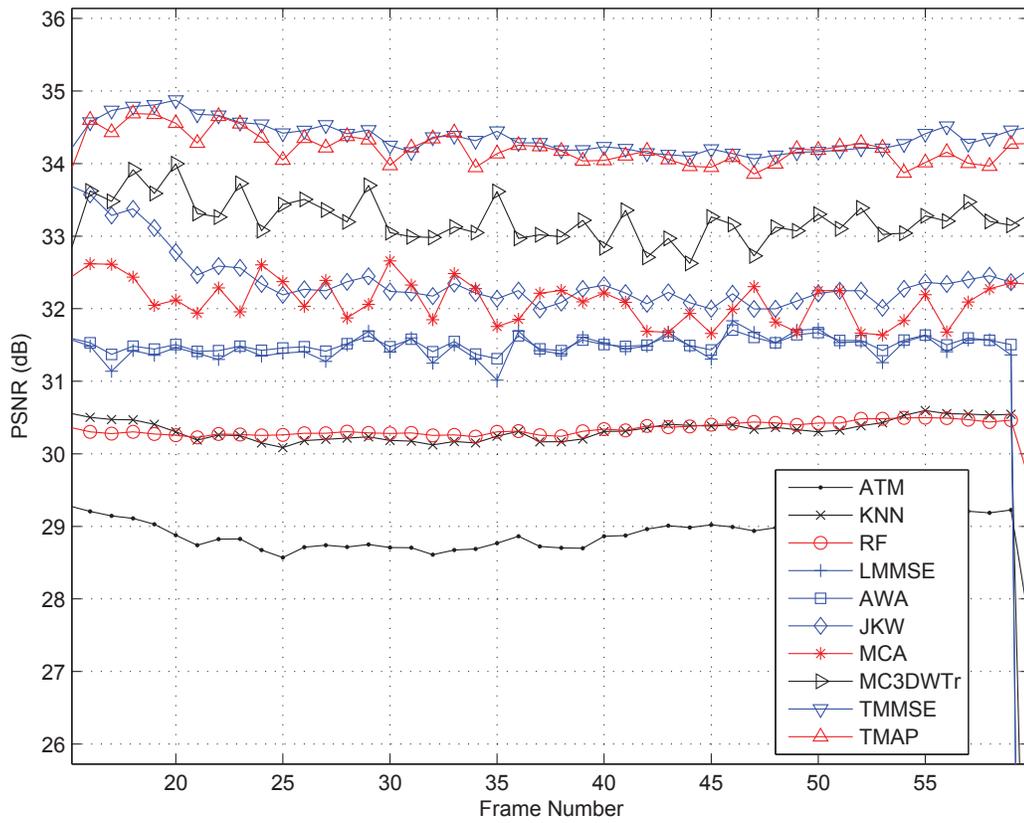


Figure 5.6: PSNR comparison of the different algorithms for *Hall* sequence having an input PSNR of 25 dB

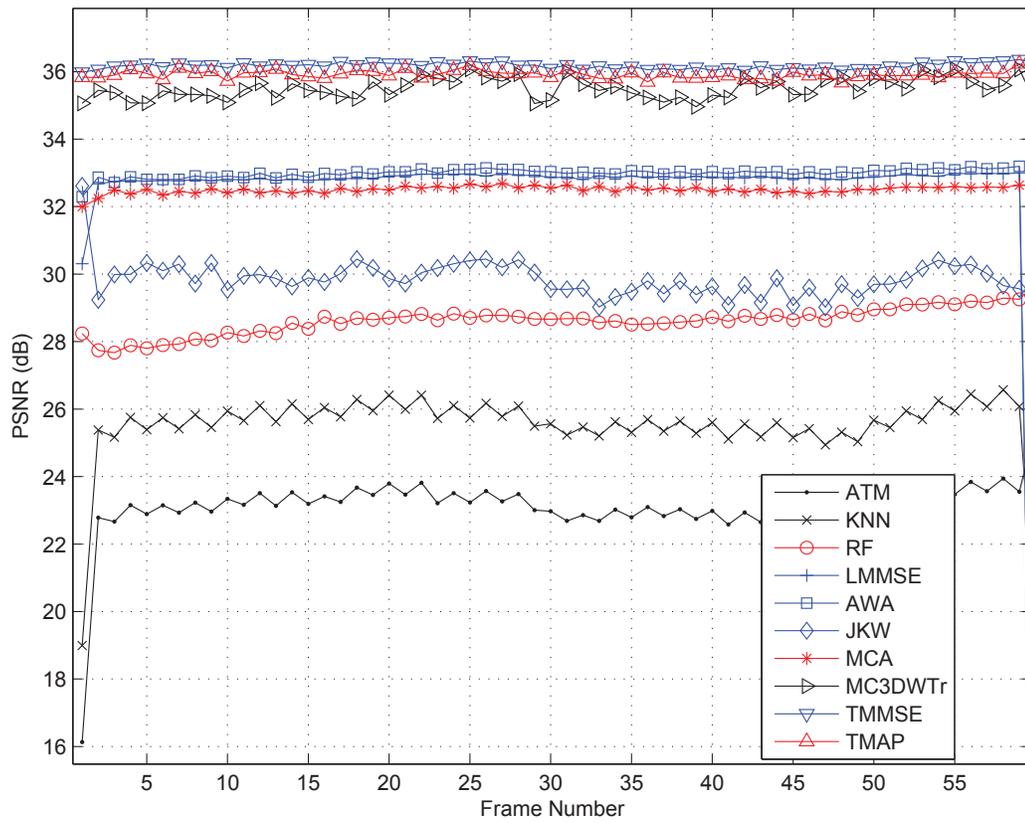


Figure 5.7: PSNR comparison of the different algorithms for *Patrol Car* sequence having an input PSNR of 30 dB

PSNRs of 20 dB, 25 dB and 30 dB, respectively. The spatial filters are not able to reduce the corrupting noise aggressively and thus the proposed filter outperforms these filters to the tune of 4-6 dB. Even compared to the wavelet domain filter (MCA), the proposed filter yields an improvement of about 0.5-3 dB. Furthermore, the variation in the output PSNRs of the proposed filters over all the processed frames is not high, which signifies that the proposed filter is able to adapt to the motion that is occurring in the video without effecting its performance. Only for the MA sequence the MC3DWTr filter marginally outperforms the proposed filters but for all other sequences the proposed filters yield a consistently better PSNR curve.

A visual comparison of the outputs generated by the different filters is presented in Fig. 5.8 using one of the frames from the PC video sequence which has an input PSNR of 25 dB. As is evident, the filters used for comparison (AWA and JKW) are unable to remove most of the noise corrupting the input frame and also cause blurring of the detail information. The performance of the proposed wavelet domain filters is superior to these filters in terms of noise removal as well as detail preservation. Among all the filters, the TMMSE filter yields the best results visually, while the performance of the TMAP filter is also very similar to the TMMSE. Both the proposed filters are able to suppress noise more aggressively than the MC3DWTr filter while preserving more detail. The visual difference is even more pronounced in the actual video as the proposed filters leads to the least amount of “flicker” like artifacts which usually result as a result of denoising the video frames spatially.

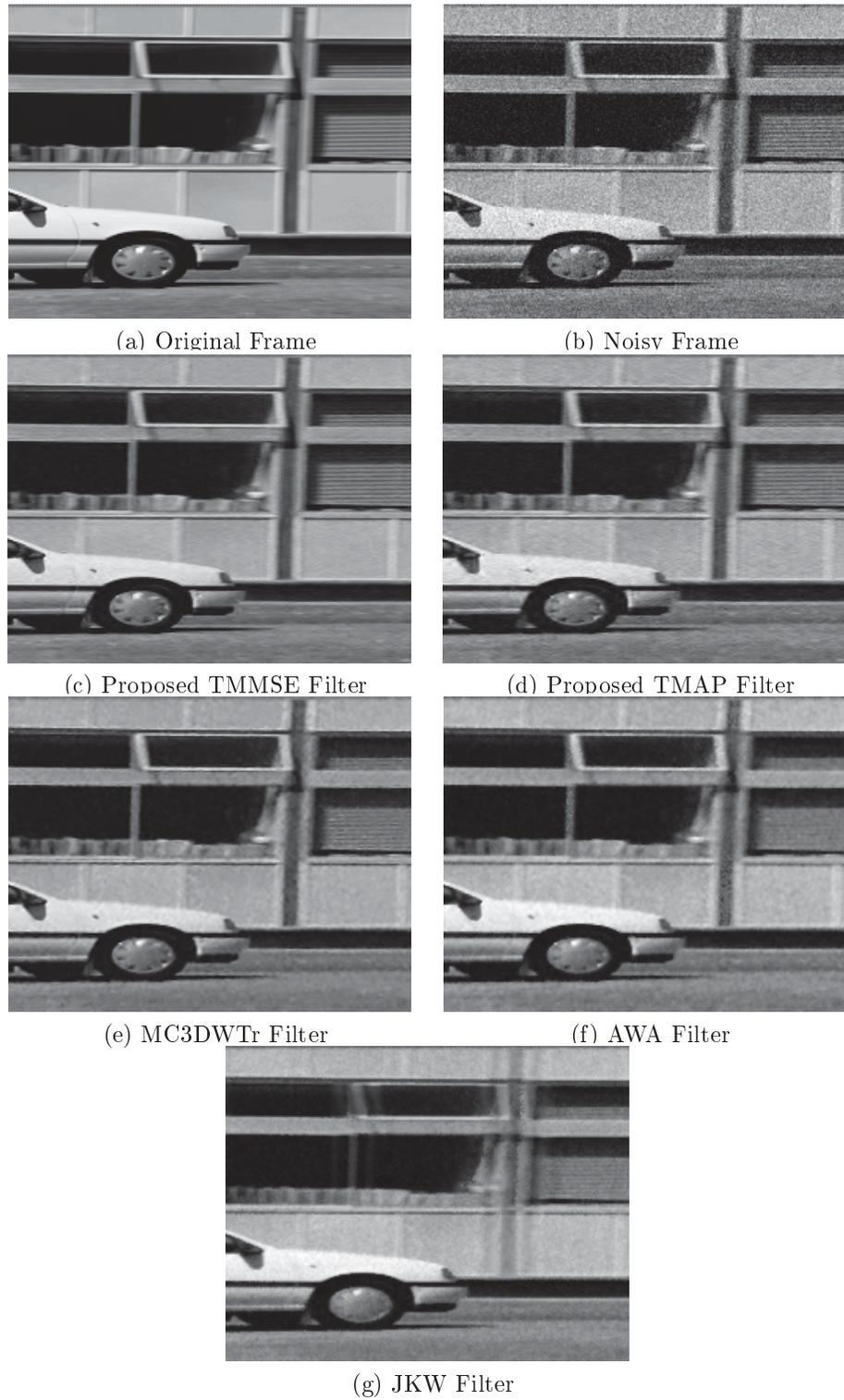


Figure 5.8: Filtered (and zoomed) outputs of one frame of the *Patrol Car* sequence for the different filters

5.6 Conclusion

In this chapter, Bayesian algorithms for video denoising in the wavelet domain have been proposed using the video modeling framework introduced in this dissertation. The fact that the spatial as well as temporal redundancies that are visible in the video sequence in the time domain also exist in the wavelet domain has been utilized for the formulation of these filters. The Bayesian filters have been developed using the minimum mean squared error as well as the maximum *a posteriori* criteria. The proposed Bayesian operators estimate the noise-free subband coefficients in the current frame, given the noisy coefficients in the current frame and the denoised estimates of the past frame. Thus, using this formulation, the magnitude of the corresponding subband coefficient in the past frame has a effect on how the shrinkage function is applied to the noisy coefficient in the current frame. This allows for a more adaptive and robust noise reduction mechanism, which can utilize both the spatial as well as the temporal redundancies within the same operator.

It has also been shown using extensive experimentation that the proposed schemes outperform several state-of-the-art video noise reduction algorithms quantitatively as well as in terms of the visual quality.

Chapter 6

Video Noise Reduction Using Adaptive Combination of Temporal and Spatial Filters

6.1 Introduction

In the previous chapters, the noise reduction of video sequence has been accomplished in the wavelet domain via the use of spatio-temporal algorithms which utilize not only the current frame subband coefficients, but also the corresponding coefficients in the neighboring frames. In order to utilize this strategy, one or more frames have to be stored in the memory along with the current frame. The larger is the temporal support that needs to be utilized, the more extensive are the memory requirements.

To counteract this problem and to keep memory requirements low while exploiting the temporal redundancies to the fullest, in this chapter we consider the development of an algorithm which filters spatial and temporal information sepa-

rately and then adaptively combines the independent estimates of the two filters based on the reliability of the motion estimates. The Kalman filter has been widely used for the filtering of temporal processes and here it is applied for the noise reduction of the video sequences in the wavelet domain via utilizing the video modeling framework presented in Chapter 3. For spatial noise reduction, the Bayesian shrinkage operator based on MMSE and presented in Section 5.3 has been used.

The organization of the chapter is as follows. The problem is briefly formulated in Section 6.2. In Section 6.3, the Kalman filter is developed which operates over the temporal motion trajectories. The adaptive combination of the estimate generated by the Kalman filter and the Bayesian MMSE processor is provided in Section 6.4 while the experimental results are discussed in detail in Section 6.5.

6.2 Problem Formulation

As discussed in Section 3.3, the subband coefficients of the noise-free video sequence frame can be expressed in terms of the corresponding subband coefficients in the past frame and the incurred motion noise, i.e.,

$$\mathbf{X}^n = \mathbf{X}^{n-1} + \mathbf{M}^{n-1}. \quad (6.1)$$

The above equation is the equivalent of a controlled stochastic process, where the process noise is similar to the observed motion noise. The motion noise has been shown to have a generalized Gaussian distribution with the model parameters ν_m and s_m . For the special case where the shape parameter is equal to 2 (i.e., $\nu_m = 2$), the motion (process) noise can be assumed to be an i.i.d. random variable

belonging to the normal distribution, such that

$$M_{ij}^n \in \mathcal{N}(0, \sigma_m^2) \quad (6.2)$$

where (i, j) denotes the spatial coordinates, n denotes the frame index and σ_m^2 is the variance of the motion noise. In this chapter, we assume this formulation of the motion noise, where it can be modeled as a normally distributed random variable.

The noisy subband coefficients for the frame n , Y_{ij}^n , which are actually observed in a captured video sequence, can be expressed as

$$Y_{ij}^n = X_{ij}^n + \epsilon_{ij}^n, \quad i = 1, \dots, L \text{ and } j = 1, \dots, K \quad (6.3)$$

where X_{ij}^n are the unknown noise-free wavelet coefficients for the frame n and ϵ_{ij}^n are i.i.d. and normally distributed random variables (measurement noise), such that

$$\epsilon_{ij} \in N(0, \sigma_\epsilon^2). \quad (6.4)$$

The estimated measurement noise variance, $\hat{\sigma}_\epsilon^2$, can be estimated as [45]

$$\hat{\sigma}_\epsilon^2 = \left(\frac{\text{median}(|HH_1|)}{0.6745} \right)^2, \quad (HH_1 \in \mathbf{Y}^n) \quad (6.5)$$

where HH_1 is the largest detail subband in the set of observed noisy subband coefficients, \mathbf{Y}^n .

Since the motion (process) noise is generated by the motion occurring in a scene while the measurement noise is generated during the digital video capture process, the motion noise and the measurement noise are assumed to be independent of each other. Thus, to estimate the variance of the motion noise, the noisy

subband coefficients are used along with the estimated noise-free coefficients from the previous frame. A noisy version of \mathbf{M}^{n-1} is obtained by subtracting $\hat{\mathbf{X}}^{n-1}$ from \mathbf{Y}^n . Using the independence of the motion noise and measurement noise, the variance of the motion noise can be estimated as

$$\hat{\sigma}_m^2 = \left(\frac{\text{median}(|HH_1|)}{0.6745} \right)^2 - \hat{\sigma}_\epsilon^2, \quad \left(HH_1 \in \left\{ \mathbf{Y}^n - \hat{\mathbf{X}}^{n-1} \right\} \right) \quad (6.6)$$

where HH_1 is the largest detail subband in the set of subband coefficients obtained by subtracting the estimated noise-free coefficients belonging to the previous frame from the observed noisy subband coefficients.

Using the system description provided by (6.1) and (6.3), this chapter addresses the problem of estimating the estimate of the noise-free subband coefficients, $\hat{\mathbf{X}}^n$, using the observed noisy coefficients, \mathbf{Y}^n .

6.3 Wavelet Domain Temporal Kalman Filter

The description of the subband coefficients in a video sequence in the form described by (6.1) and (6.3) makes the estimation of the noise-free subband coefficients an ideally suited problem for time recursive filtering. In the time domain, Dugad *et al.* [39] have proposed a Kalman filter which operates on the pixel values in the temporal direction for the estimation of the denoised pixels. Based on the proposed modeling of the subband coefficients, it is also possible to formulate such a time recursive Kalman filter in the wavelet domain also. Such a filter works purely in the temporal direction along the motion trajectories in the video sequence and thus necessitates the compensation of motion prior to the application of the filter in order to find the subband coefficients corresponding to the current one in the past frame.

The Kalman filter has to be initialized before it can process the first frame. To perform this initialization, the spatial estimate of the wavelet coefficient of the first frame (which is obtained as described in the next section) is used and the initial error is initialized as the variance of the measurement noise, $\hat{\sigma}_\epsilon^2$.

The Kalman filter first performs the time update step wherein it estimates the value of the current wavelet coefficient using the past available data along the motion trajectory and then performs the measurement update step, where it updates these estimates based on the observed value of the current wavelet coefficient.

Initialization:

$$\hat{X}_{ij}^{0|0} = \text{Spatial Estimate}, \hat{X}_{ij}^1 \quad (6.7)$$

$$P_{ij}^{0|0} = \hat{\sigma}_\epsilon^2 \quad (6.8)$$

Time Update:

$$\hat{X}_{ij}^{n|n-1} = \hat{X}_{ij}^{n-1|n-1} \quad (6.9)$$

$$P_{ij}^{n|n-1} = P_{ij}^{n-1|n-1} + \hat{\sigma}_m^2 \quad (6.10)$$

Measurement Update:

$$\hat{X}_{ij}^{n|n} = \hat{X}_{ij}^{n|n-1} + \frac{P_{ij}^{n|n-1}}{P_{ij}^{n|n-1} + \hat{\sigma}_\epsilon^2} (Y_{ij}^n - \hat{X}_{ij}^{n|n-1}) \quad (6.11)$$

$$P_{ij}^{n|n} = P_{ij}^{n|n-1} \left(1 - \frac{P_{ij}^{n|n-1}}{P_{ij}^{n|n-1} + \hat{\sigma}_\epsilon^2} \right) \quad (6.12)$$

$\hat{X}_{ij}^{n|n}$ is the required estimate generated by the Kalman filter and for a wavelet coefficient at location (i, j) in the frame n , and for future processing, can be denoted as $\hat{X}k_{ij}^n$.

Thus, the Kalman filter updates the estimation error recursively and based on the updated error variance, estimates the noise-free subband coefficient in the frame n . Based on the magnitude of the estimated motion noise, the Kalman filter either allows or dampens the temporal smoothing of the observed noisy coefficient, thus, making it an adaptive recursive temporal filter.

6.4 Sequential Combination of Spatial and Temporal Estimates

The process described in the previous section provides an estimate of the noise-free subband coefficients, and is based solely on the temporal behavior of the video sequence. It is also possible to obtain a spatial estimate of the same subband coefficients using the Bayesian MMSE processor described in Section 5.3. This operator has a limited temporal support of one frame and thus, provides an independent estimate of the subband coefficients, which can be expressed as

$$\hat{X}_{s_{ij}}^n = \text{TMMSE} \left(Y_{ij}^n, \hat{X}_{ij}^{n-1} \right) \quad (6.13)$$

where Y_{ij}^n and \hat{X}_{ij}^{n-1} are the noisy subband coefficients in the current frame and the corresponding estimate of the noise-free coefficients in the past frame.

Having obtained the two estimates (spatial and temporal) of the same subband coefficient, these two estimates need to be combined in order to obtain the robust spatio-temporal estimate of the noise-free subband coefficient in the current frame.

Since the Kalman filter acts along the motion trajectories, it is sensitive to motion estimation and in case a lot of scene motion is observed or if the estimated motion compensation is not reliable, the estimate produced by the Kalman filter should not be used to generate the final estimate. On the other hand, in video sequences having no or very small motion, the Kalman filter performs as good as the spatial filter and an equal weight for both the estimates is desired. Thus, depending upon the estimated motion noise variance, the two estimates can be combined to obtain the final estimate, \hat{X}_{ij}^n , as follows:

$$\hat{X}_{ij} = \begin{cases} (1-w)\hat{X}s_{ij}^n + w\hat{X}k_{ij}^n, & \text{if } \hat{\sigma}_m^2 < \hat{\sigma}_e^2 \\ \hat{X}s_{ij}^n, & \text{if } \hat{\sigma}_m^2 > \hat{\sigma}_e^2 \end{cases} \quad (6.14)$$

where the weighting function chosen is an exponential function, which is proportional to the ratio of the variances of the motion and the measurement noises and can be described as

$$w = \frac{1}{2} \exp\left(-\frac{\hat{\sigma}_m^2}{10}\right). \quad (6.15)$$

If the motion noise variance is low, then the estimated temporal estimate is reliable and thus, both the spatial and temporal are utilized to obtain the final estimate. However, if the motion noise variance is high due to the unreliable motion estimates, then only the the spatial filter is utilized in the final estimate.

The output of the inverse wavelet transform operation on the denoised coefficients in the current frame yields the denoised time domain pixel values as

$$\hat{\mathbf{f}}^n = \text{InverseDWT}(\hat{\mathbf{X}}^n) \quad (6.16)$$

6.5 Experimental Results and Discussion

The performance of the proposed algorithm which is based on the adaptive combination of the spatially and temporally filtered video sequences, JKMMSE, is evaluated using the following state-of-the-art algorithms:

- in the time-domain: 3D Rational filter (RF) [26], 3D KNN filter [24], 3D α -trimmed mean (ATM) filter [24], 3D LMMSE filter [35], the adapted weighted averaging (AWA) filter [35] and the joint Kalman and Wiener filter (JKW) [39].
- in the wavelet-domain: multiple class averaging (MCA) algorithm [29] and MC-3DWTr [58].

As metrics for the evaluation of the different algorithms, as described in Section 4.6.1, peak signal-to-noise ratio (PSNR) and visual quality are used.

Five different test sequences, viz. *Hall* (Ha), *Miss America* (MA), *Foreman* (Fo), *Patrol Car* (PC) and *Train* (Tr), corrupted by three different AWGN powers are used for evaluation. The sequences have varying amount of motion (both local and global) as well as detail. The input PSNRs of the corrupted sequences are 20 dB, 25 dB and 30 dB. The necessary motion compensation of the past frames is accomplished in the wavelet domain using the half-pixel level block matching technique with a block size of 8x8. The motion vectors are estimated using the past denoised frame and the current noisy frame; the estimated motion vectors are then used for compensating motion in the past and the future noisy sequences. The Symmlet-8 wavelet is used to decompose the input frames upto 4 dyadic levels.

Table 6.1 presents the average PSNR values for the different algorithmic outputs with the sequences having PSNRs of 20 dB, 25 dB and 30 dB as input. The

	ATM	KNN	RF	MCA	LMMSE	AWA	JKW	MC3DWT _T	JKMMSE
Input PSNR = 20 dB									
MA	30.37	29.0	29.29	31.21	30.96	29.01	30.65	35.35	34.85
Ha	27.66	27.29	27.31	28.09	27.35	27.75	28.68	29.57	31.61
Fo	27.96	27.75	28.11	28.07	26.14	28.09	27.64	29.85	31.45
PC	22.94	24.14	25.97	26.15	25.5	25.6	24.89	27.99	29.24
Tr	22.38	23.89	24.27	25.53	25.3	25.22	25.59	26.72	27.85
Average	26.25	26.41	26.95	27.81	27.05	27.15	27.48	29.89	31.0
Input PSNR = 25 dB									
MA	33.61	33.45	34.18	34.92	34.57	33.36	32.22	37.32	36.82
Ha	28.74	29.98	30.34	32.04	30.86	31.07	32.37	32.92	34.71
Fo	30.58	30.37	31.66	31.44	30.13	29.85	29.42	32.59	33.93
PC	23.06	25.18	27.93	29.95	28.83	29.08	27.41	31.31	32.46
Tr	22.9	24.8	25.43	29.25	28.74	28.6	28.25	30.08	31.38
Average	27.78	28.75	29.91	31.52	30.63	30.39	29.97	32.84	33.86
Input PSNR = 30 dB									
MA	37.69	37.19	37.53	37.78	37.66	37.3	37.25	38.97	38.98
Ha	29.2	31.38	31.64	35.13	34.66	34.97	35.11	36.05	37.57
Fo	31.25	32.00	33.44	34.65	33.95	31.15	31.72	35.41	36.63
PC	23.06	25.48	28.62	33.2	32.38	32.55	29.85	34.98	36.21
Tr	22.8	25.15	25.8	32.35	32.44	31.97	30.57	33.81	34.92
Average	28.79	30.24	31.41	34.61	34.21	33.59	32.90	35.85	36.86

Table 6.1: PSNR comparison of the proposed algorithm, JKMMSE, with several others for the corrupted input test sequences with input PSNRs of 20 dB, 25 dB and 30 dB, resp.

proposed filter, JKMMSE, yields an average PSNR gain in the range of 7 to 15 dB over the corrupted sequences and gains of 1-7.5 dB over the compared schemes. The sequences in which the amount of motion is limited have a higher increase in the PSNR even when compared to the TMMSE filter of the previous chapter. This is so because with limited motion, the Kalman filter produces valid estimates when using only the temporal data and thus, when combined with the spatially filtered outputs with an almost equal weight, the resultant PSNR is better than those of the two filters individually. This increase is to the tune of 0.3-0.75 dB depending upon the amount of motion in the sequence. The amount of increase observed is also dependent upon the level of detail in the sequence and for sequences with a large quantity of fine detail (such as Tr), the overall increase in the PSNR is reduced. Since the Kalman filter acts along the motion compensated trajectories, when combined with the spatial filter, it is still able to provide an overall improvement in the PSNR even for sequence with significant local and global motion. Among all the techniques used for comparison, the performance of the MC3DWTr is the closest to the proposed algorithm, with the proposed algorithm outperforming MC3DWTr by ~ 1 dB on average.

Figures 6.1, 6.2 and 6.3 show the PSNR curves obtained using the different algorithms when processing Ha, PC and MA sequences corrupted to have input PSNRs of 20 dB, 25 dB and 30 dB, respectively. The proposed filter outperforms the spatial filters to the tune of 4-6 dB and even compared to the wavelet domain filter (MCA), the proposed filter yields an improvement of about 0.5-3 dB. Furthermore, even for the sequence containing significant motion (PC), the overall performance of the combined filter is maintained over all the frames in the sequence. Even when compared to the state-of-the-art MC3DWTr filter, the proposed filter outperforms consistently in terms of PSNR.

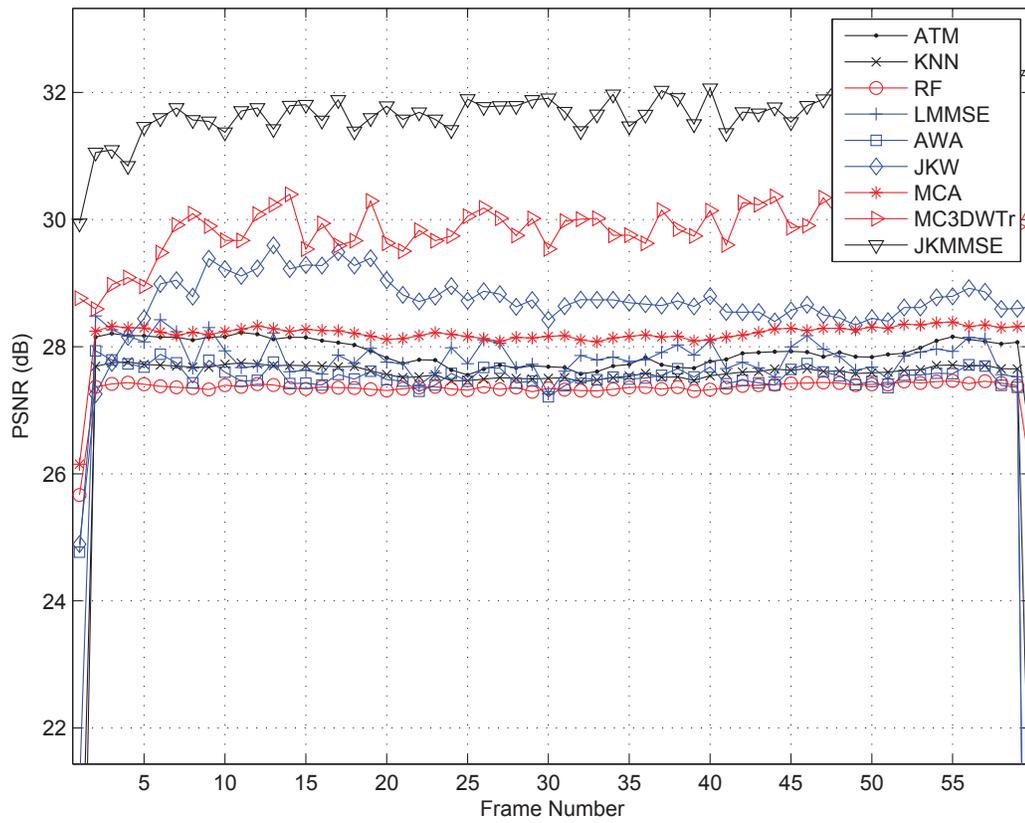


Figure 6.1: PSNR comparison of the different algorithms for *Hall* sequence having an input PSNR of 20 dB

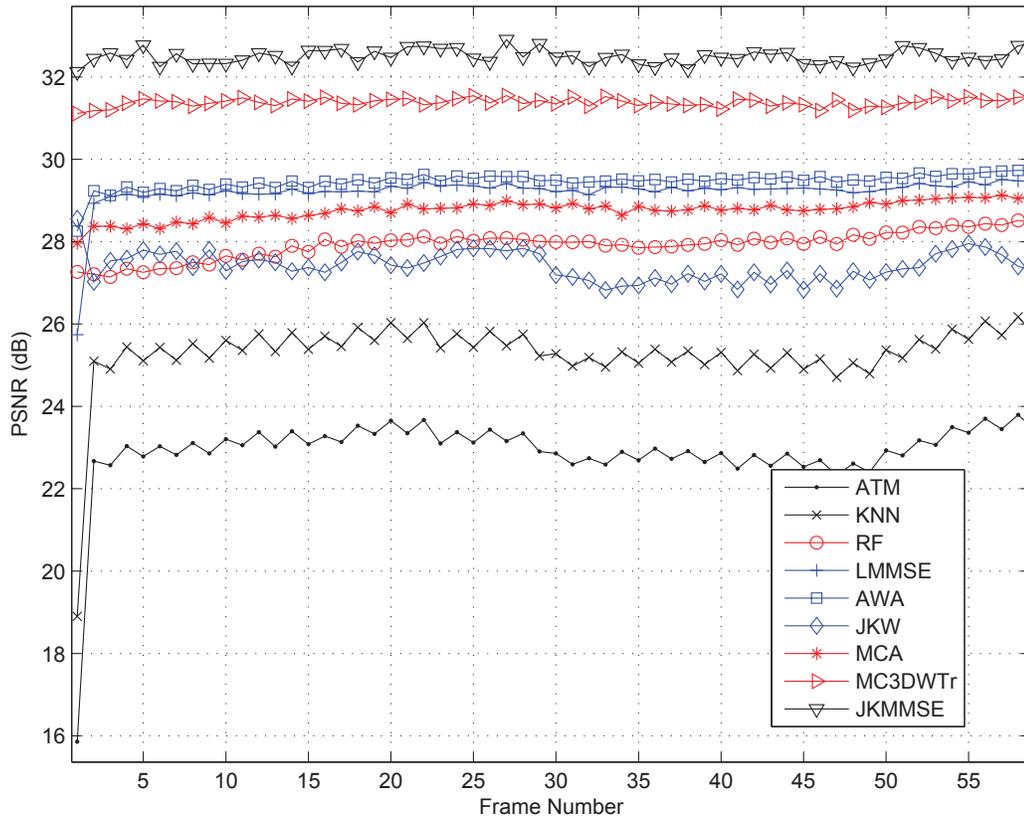


Figure 6.2: PSNR comparison of the different algorithms for *Patrol Car* sequence having an input PSNR of 25 dB

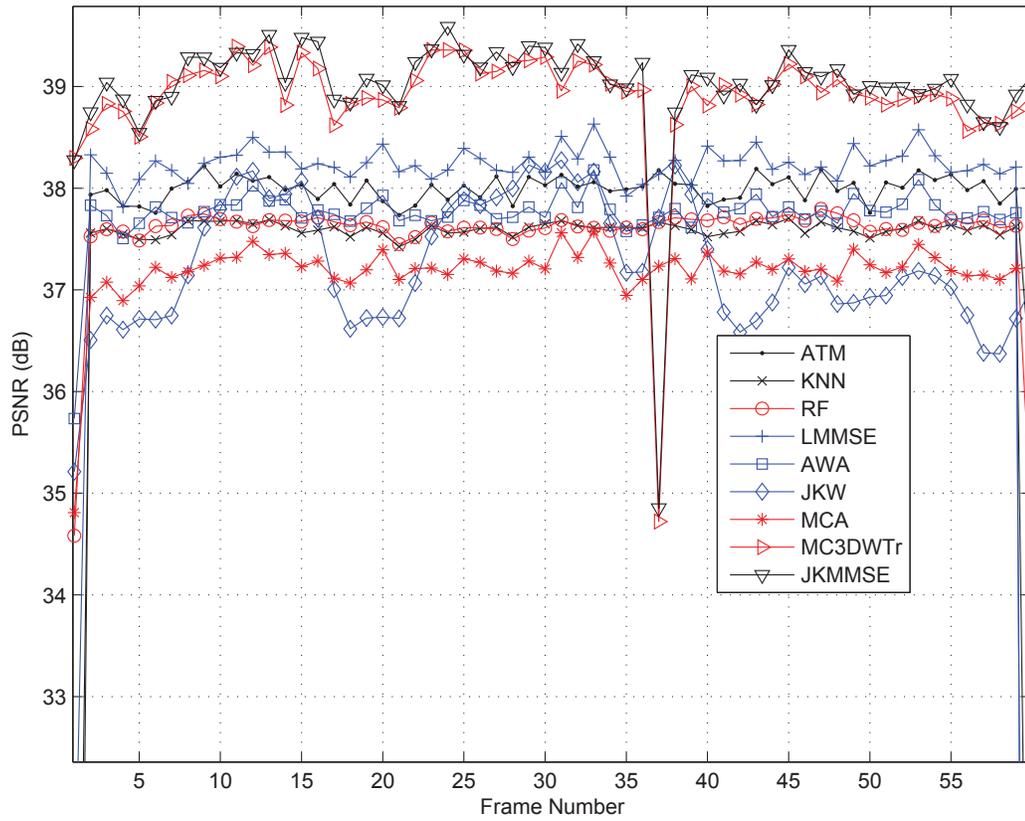


Figure 6.3: PSNR comparison of the different algorithms for *Miss America* sequence having an input PSNR of 30 dB

A visual comparison of the outputs generated by the different filters is presented in Fig. 6.4 using one of the frames from the MA video sequence which has an input PSNR of 20 dB. As is evident, the filters used for comparison (ATM, LMMSE and MCA) are unable to remove most of the noise corrupting the input frame and also cause significant blurring. The performance of the proposed JK-MMSE filter is superior to these filters in terms of noise removal as well as detail preservation. The MC3DWTr filter comes the closest to the proposed filter in terms of noise suppression when compared visually. Since the Kalman filter acts only temporally, it is able to suppress any flicker artifacts which are introduced with spatial-only filtering of the video sequences and when combined with the spatio-temporal TMMSE filter, the overall result is a visually more pleasing video sequence.

6.6 Conclusion

This chapter discussed a novel wavelet domain noise reduction scheme, where the spatial and temporal noise reduction was performed separately based on the video modeling framework introduced in this dissertation. The individually filtered results were then combined adaptively to generate a single noise-free estimate of the original subband coefficient of the video sequence. For the noise reduction in the temporal direction, a Kalman filter has been used and low-complexity techniques have been suggested for the computation of the process and measurement noises. The spatio noise reduction has been achieved using the adaptive TMMSE operator described in the previous chapter. The results obtained from both the filters are then combined based on the amount of motion observed in the sequence. If the motion is sufficiently compensated, then the temporal filter is weighted almost

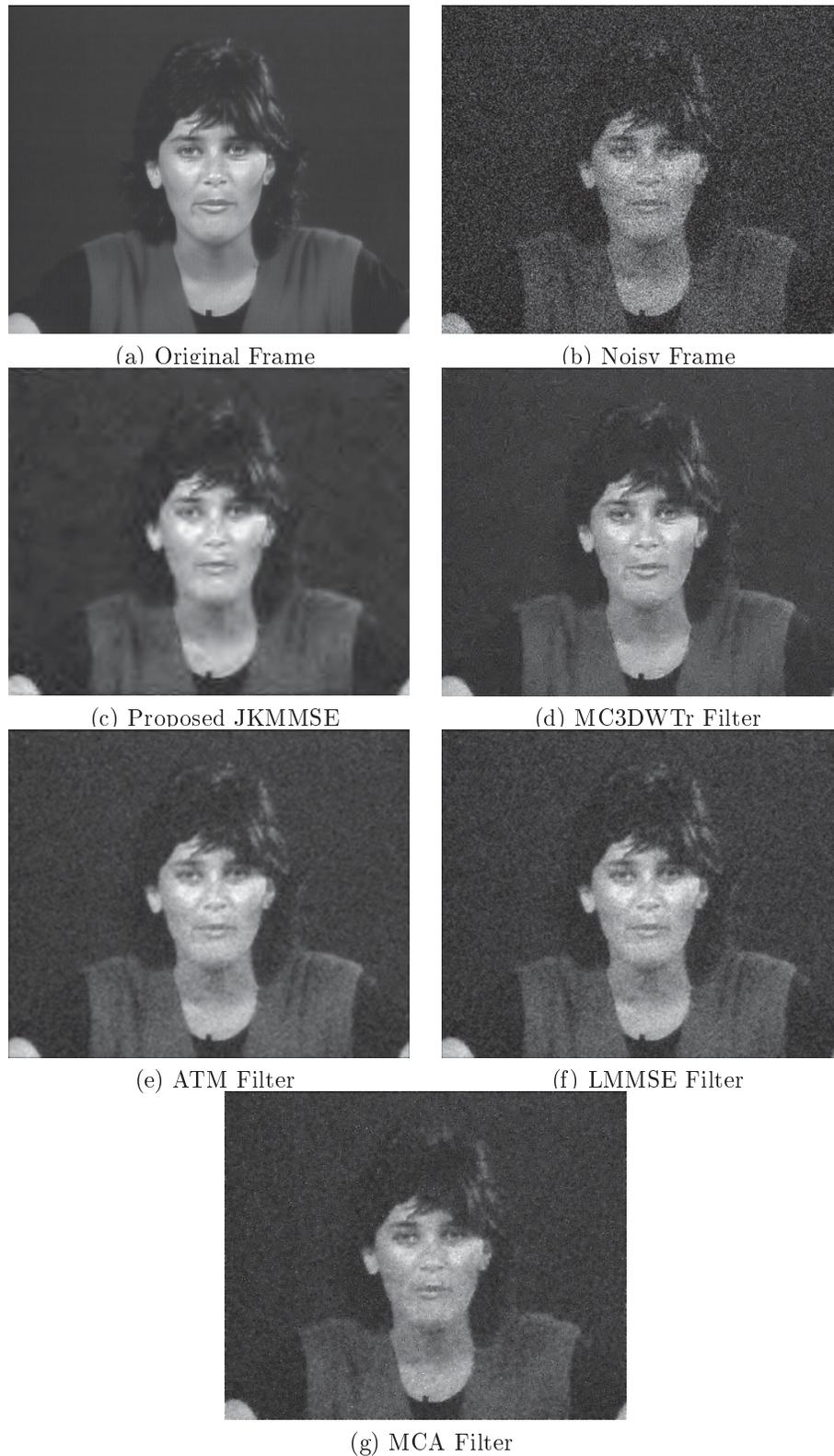


Figure 6.4: Filtered outputs of the 17th frame of the *Miss America* sequence for the different algorithms

equally with the spatial filter to generate the final output. On the other hand, if a significant amount of motion noise is detected, then the final result is more heavily dependent on the spatially filtered output.

Extensive simulations have been performed and the presented results when compared with several state-of-the-art algorithms in literature confirm the superior performance of the proposed technique.

Chapter 7

Motion Adaptive Video Denoising Based on Bivariate Shrinkage of Wavelet Coefficients

7.1 Introduction

The previous chapters have addressed the problem of noise reduction by formulating algorithms that operate entirely in the wavelet domain. The temporal redundancies that exist in the subband coefficients belonging to successive frames have been exploited in a variety of ways, including decorrelation, condition Bayesian filtering and Kalman filtering. The interscale correlation, on the other hand, has been addressed only in Chapter 4 in a heuristic spatial thresholding mechanism. It is also possible to exploit the interscale redundancies in a more structured Bayesian framework and develop an operator which adapts to these spatial correlations.

Thus, in this chapter, a new wavelet-based video noise reduction algorithm is discussed, where the inter-scale dependencies in the subband representation of

each frame are modeled using a non-Gaussian bivariate distribution. The parameters for this bivariate distribution are estimated adaptively using the local correlations that exist between neighboring coefficients within each subband. Based on this bivariate distribution, a shrinkage function is developed using the *maximum a posteriori* (MAP) rule. To improve the performance of the filter, information from the previous frames is also incorporated in the shrinkage function by detecting the motion between the corresponding subband coefficients in successive frames.

The chapter is organized as follows. The formulation of the problem and the modeling of the video sequence is presented in Section 7.2, while the proposed MAP estimation is developed in Section 7.3. Section 7.4 discusses the simulation results and performance comparison of the proposed algorithm with several other state-of-the-art algorithms for video noise reduction..

7.2 Problem Formulation and Bivariate Modeling

Similar to previous chapters, the denoising of a video sequence corrupted by additive white Gaussian noise is considered:

$$\mathbf{g}^n = \mathbf{f}^n + \eta^n \quad (7.1)$$

where \mathbf{g}^n and \mathbf{f}^n are the n^{th} frames or fields of the noisy and the noise-free sequences, respectively. The video noise, η^n , is considered to be zero-mean AWGN having a variance σ_η^2 . Given the observed noisy frame, \mathbf{g}^n , an estimate of the original noise-free video sequence, \mathbf{f}^n , is sought here. The input noisy frame is transformed to the wavelet domain and the problem can be formulated as

$$\mathbf{Y}^n = \mathbf{X}^n + \epsilon^n \quad (7.2)$$

where \mathbf{Y}^n and \mathbf{X}^n are the sets of noisy and noise-free wavelet coefficients, respectively. ϵ^n is the set of corrupting zero-mean AWGN coefficients having a variance σ_ϵ^2 , which can be estimated from the noisy wavelet coefficients in the diagonal detail subband via a use of the robust estimator [45], $\hat{\sigma}_\epsilon^2 = \left(\frac{\text{median}(|HH_1|)}{0.6745} \right)^2$.

The distribution of wavelet coefficients of video frames is sharp-peaked and heavy-tailed and can be modeled using the generalized Gaussian (GG) distribution (see Chapter 3). However, it is difficult to model the statistical dependencies between wavelet coefficients using such marginal priors. As shown in Chapter 4, there exist dependencies between the subband wavelet coefficients of a video frame and the corresponding parent coefficients (in the adjacent coarser scale). Let x_1^n denote a subband coefficient for the noise-free frame, n , and its parent at the next coarser level be denoted as x_2^n . Then

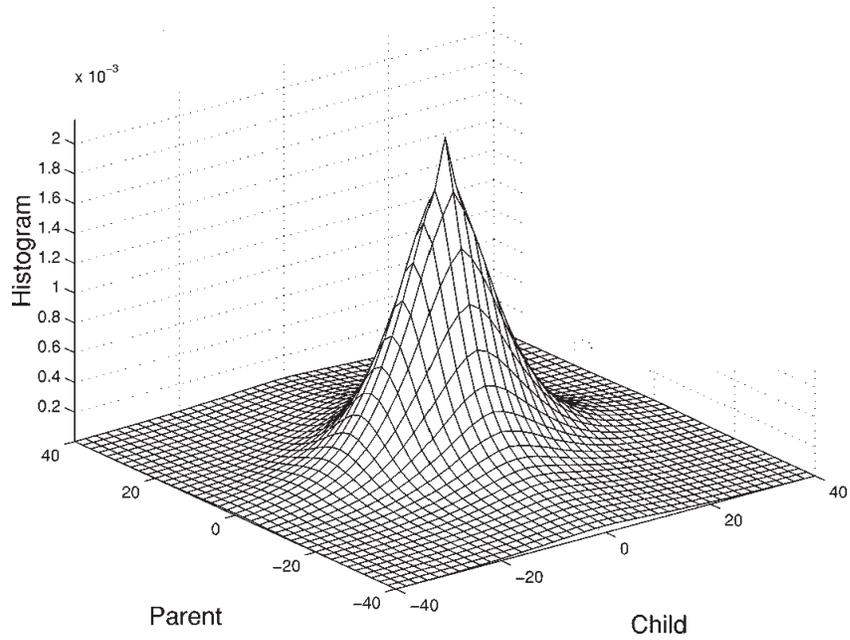
$$\begin{aligned} y_1^n &= x_1^n + \epsilon_1^n \\ y_2^n &= x_2^n + \epsilon_2^n \end{aligned} \tag{7.3}$$

where y_1^n and y_2^n are the noisy observations of x_1^n and x_2^n , and ϵ_1^n and ϵ_2^n are the noise samples.

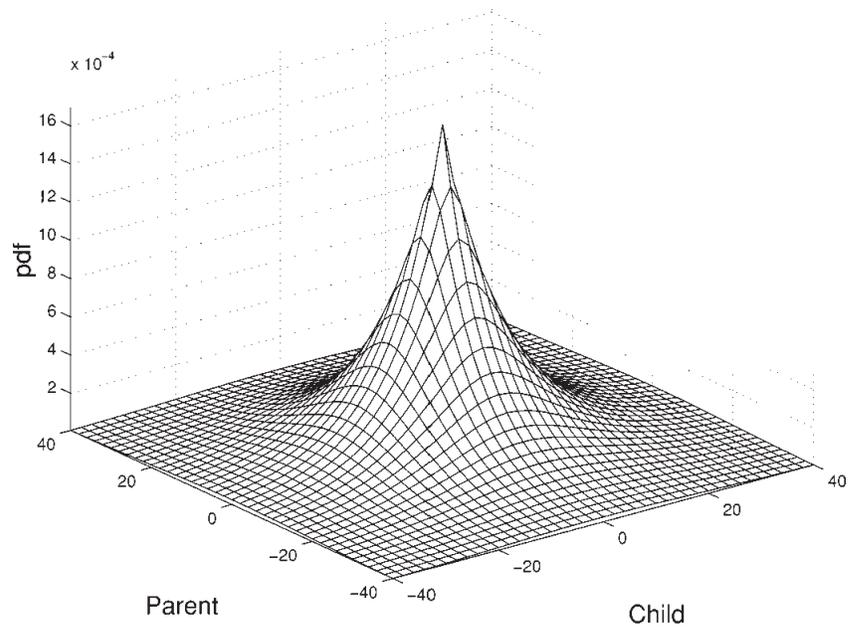
Figure (7.1) (a) shows the joint histogram obtained using the noise-free coefficients belonging to the HL_1 and HL_2 subbands of a noise-free frame from the *Trevor* video sequence. This joint distribution has been modeled using a non-Gaussian distribution [67] which can be described as

$$p_{x_1, x_2}(x_1, x_2) = \frac{3}{2\pi\sigma_{xx}^2} \exp\left(-\frac{\sqrt{3}}{\sigma_{xx}} \sqrt{x_1^2 + x_2^2}\right), \tag{7.4}$$

where $p_{x_1, x_2}(x_1, x_2)$ is the joint pdf of the random variables x_1 and x_2 and σ_{xx}^2 is the variance of the distribution. Figure (7.1) (b) shows the plot of the distribu-



(a)



(b)

Figure 7.1: (a) Empirical histogram obtained using the subband coefficients; (b) Histogram obtained using (7.4)

tion described by (7.4) which is used to model the bivariate empirical histogram of the parent and child subband coefficients. As is evident, the used bivariate non-Gaussian distribution is able to model the empirical histogram very closely. The sharp peak of the empirical histogram at the mean value (0) and the slowly tapering tail is also reflected in the plot of the distribution used for modeling the empirical histogram.

7.3 Adaptive Bivariate Shrinkage

In this section, the Bayesian processor based on the MAP estimation to estimate the noise-free wavelet coefficients, \hat{x}_1^n , corresponding to frame n , is described. The noisy subband coefficients in the current frame, y_1^n , along with their corresponding parents, y_2^n , are used in the estimation process. We can write

$$\mathbf{y}^n = \mathbf{x}^n + \epsilon^n \quad (7.5)$$

where $\mathbf{y}^n = \{y_1^n, y_2^n\}$, $\mathbf{x}^n = \{x_1^n, x_2^n\}$ and $\epsilon^n = \{\epsilon_1^n, \epsilon_2^n\}$.

The MAP estimate of the noise-free subband coefficient can be obtained as

$$\hat{\mathbf{x}}^n(\mathbf{y}^n) = \arg \max_{\mathbf{x}^n} p_{\mathbf{x}^n|\mathbf{y}^n}(\mathbf{x}^n|\mathbf{y}^n). \quad (7.6)$$

This equation can be manipulated to be written as

$$\hat{\mathbf{x}}^n(\mathbf{y}^n) = \arg \max_{\mathbf{x}^n} [p_{\epsilon^n}(\mathbf{y}^n - \mathbf{x}^n)p_{\mathbf{x}^n}(\mathbf{x}^n)]. \quad (7.7)$$

Since the noise is i.i.d. and Gaussian,

$$p_{\epsilon^n}(\epsilon^n) = \frac{1}{2\pi\sigma_\epsilon^2} \exp\left(-\frac{(\epsilon_1^n)^2 + (\epsilon_2^n)^2}{2\sigma_\epsilon^2}\right) \quad (7.8)$$

Solving (7.7) is equivalent to solving

$$\hat{\mathbf{x}}^n(\mathbf{y}^n) = \arg \max_{\mathbf{x}^n} [\ln(p_{\epsilon^n}(\mathbf{y}^n - \mathbf{x}^n)) + \ln(p_{\mathbf{x}^n}(\mathbf{x}^n))]. \quad (7.9)$$

If we define $f(\mathbf{x}^n) = \ln(p_{\mathbf{x}^n}(\mathbf{x}^n))$, then (7.9) can be solved by solving the following two equations together:

$$\begin{aligned} \frac{y_1^n - \hat{x}_1^n}{\sigma_\epsilon^2} + f_1(\hat{\mathbf{x}}^n) &= 0 \\ \frac{y_2^n - \hat{x}_2^n}{\sigma_\epsilon^2} + f_2(\hat{\mathbf{x}}^n) &= 0 \end{aligned} \quad (7.10)$$

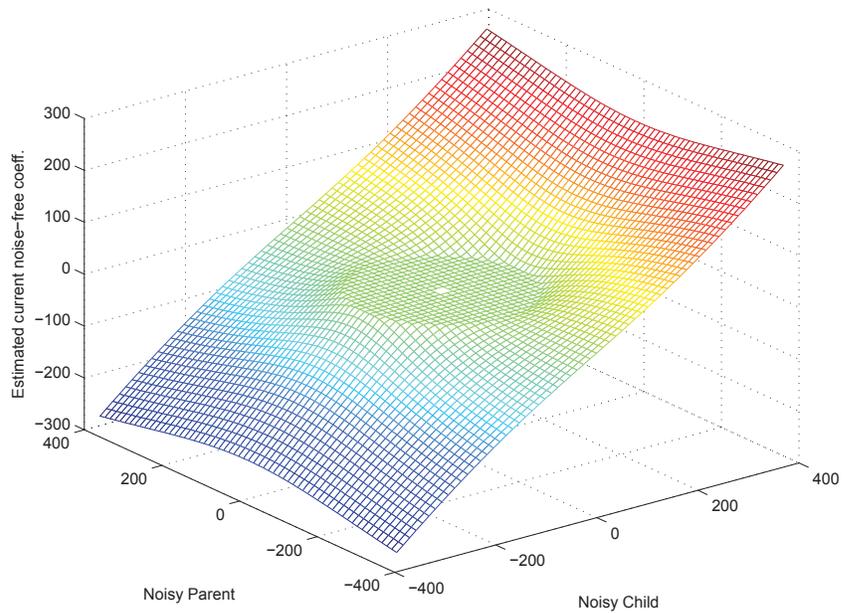
where $f_1(\hat{\mathbf{x}}^n)$ and $f_2(\hat{\mathbf{x}}^n)$ are the derivatives of $f(\mathbf{x}^n)$ with respect to x_1^n and x_2^n , respectively. Solving (7.10), the joint shrinkage rule can be expressed as [67]

$$\hat{x}_1^n = \frac{\left(\sqrt{(y_1^n)^2 + (y_2^n)^2} - \frac{\sqrt{3}\sigma_\epsilon^2}{\sigma_{xx}}\right)_+}{\sqrt{(y_1^n)^2 + (y_2^n)^2}} y_1 \quad (7.11)$$

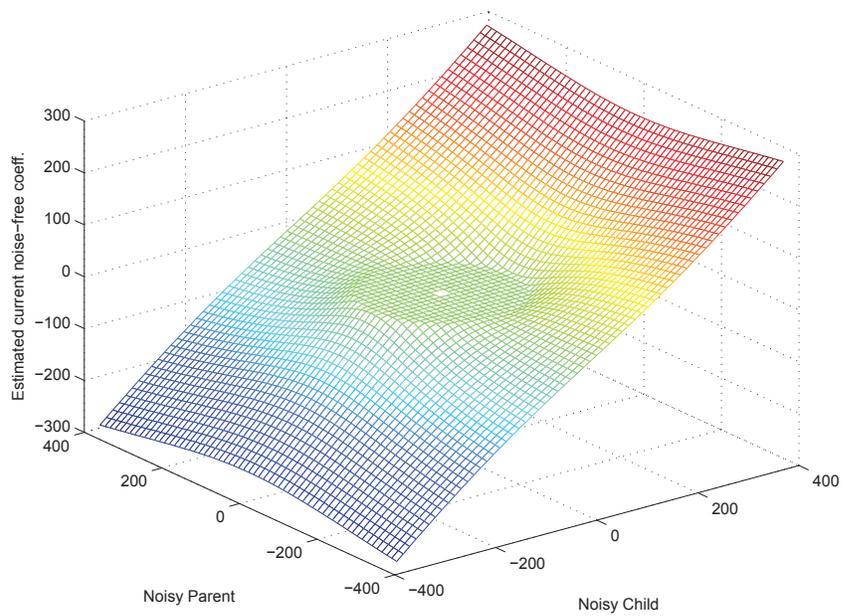
where the $(\cdot)_+$ signifies that only the positive values are considered, while the negative values are put to zero.

Figure 7.2 shows two plots of this bivariate function for different combinations of σ_{xx} and σ_ϵ . The shrinkage function has a circular dead-zone where the estimated value of the noise-free coefficients is zero. This shrinkage function maps all the small subband coefficients to zero while smoothly shrinking the others based on the values of the noisy subband coefficient and its corresponding parent.

To estimate σ_{xx} , the immediate neighborhood around the noisy subband coefficient y_1^n is utilized. For this estimation of σ_{xx} , a coefficient set \mathcal{S}^n is defined,



(a) Bivariate shrinkage function ($\sigma_\epsilon = 30$, $\sigma_{xx} = 10$)



(b) Bivariate shrinkage function ($\sigma_\epsilon = 20$, $\sigma_{xx} = 5$)

Figure 7.2: Bivariate shrinkage function obtained using the MAP estimation

which includes the noisy coefficients in the 5x5 neighborhood centered at y_1^n . Since the input is a video sequence, it is also possible to utilize the information available in the neighborhood frames, i.e., \mathbf{Y}^{n-1} and \mathbf{Y}^{n+1} . However, the motion occurring in the video might cause the statistics of the coefficients in \mathbf{Y}^{n-1} and \mathbf{Y}^{n+1} , which correspond to y_1^n , to change. Even if the motion in the video sequence is estimated and then explicitly compensated, the past and the future subband coefficients might not possess similar statistics to y_1^n . Thus, in order to detect the motion that has occurred in the region of interest between any pair of frames n and $n-1$, the neighborhoods \mathcal{S}^n and \mathcal{S}^{n-1} around each of the corresponding noisy coefficients y_1^n and y_1^{n-1} are considered. The mean absolute difference between the two regions is then computed as

$$\text{MAD}(n, n-1) = \frac{1}{25} \sum_{\substack{y_{1i}^n \in \mathcal{S}^n \\ y_{1i}^{n-1} \in \mathcal{S}^{n-1}}} |y_{1i}^n - y_{1i}^{n-1}|. \quad (7.12)$$

If the estimated difference is caused only due to the noise present in the video signal, i.e., if $\text{MAD}(n, n-1) < \sigma_\epsilon$, then the coefficients in neighborhood \mathcal{S}^{n-1} are also considered along with the coefficients in \mathcal{S}^n in order to estimate σ_{xx} . Similarly, the motion between frames n and $n+1$ is also checked and depending upon the amount of motion, the subband coefficients belonging to the set \mathcal{S}^{n+1} are included in the estimation of σ_{xx} .

If \mathcal{S} denotes the total neighborhood which includes all the corresponding coefficients in the frames n , $n-1$ and $n+1$ where significant motion has not occurred,

then σ_{xx} can be estimated as

$$\hat{\sigma}_{xx} = \sqrt{\left(\frac{1}{N_S} \sum_{y_{1i} \in \mathcal{S}} y_{1i}^2 - \sigma_\epsilon^2\right)_+} \quad (7.13)$$

where N_S are the total number of coefficients in the neighborhood \mathcal{S} .

The output of the inverse wavelet transform operation on the denoised coefficients in the current frame yields the denoised time domain pixel values as

$$\hat{\mathbf{f}}^n = \text{InverseDWT}(\hat{\mathbf{X}}^n). \quad (7.14)$$

7.4 Experimental Results and Discussion

The performance evaluation of the proposed spatio-temporally adaptive algorithm which is based on bivariate modeling of the subband coefficients and maximum *a posteriori* estimation, TBMAP, is presented in this section. Five different test sequences, viz. *Hall* (Ha), *Miss America* (MA), *Foreman* (Fo), *Patrol Car* (PC) and *Train* (Tr), corrupted by three different AWGN powers are used for evaluation. The sequences have varying amount of motion (both local and global) as well as detail. The input PSNRs of the corrupted sequences are 20 dB, 25 dB and 30 dB. The necessary motion compensation of the past frames is accomplished in the wavelet domain using the half-pixel level block matching technique with a block size of 8x8. The motion vectors are estimated using the past denoised frame and the current noisy frame; the estimated motion vectors are then used for compensating motion in the past and the future noisy sequences. The Symmlet-8 wavelet is used to decompose the input frames upto 4 dyadic levels.

The performance is compared with the 3D Rational filter (RF) [26], 3D KNN filter [24], 3D α -trimmed mean (ATM) filter [24], 3D LMMSE filter [35], the

adapted weighted averaging (AWA) filter [35] and the joint Kalman and Wiener filter (JKW) [39] (operating in time domain) and multiple class averaging (MCA) algorithm [29] and MC3DWTr [58] filter which operate in the wavelet domain.

Peak signal-to-noise ratio (PSNR) and visual quality are used as metrics for the evaluation of the different algorithms.

Table 7.1 presents the average PSNR values for the different algorithmic outputs with the sequences having PSNRs of 20 dB, 25 dB and 30 dB as input. The proposed filter, TBMAP, yields an average PSNR gain in the range of 7 to 14 dB over the corrupted sequences and gains of 0.4-6.75 dB over the compared schemes. The proposed filter is able to adapt to the different levels of detail in the video sequence as well as the different degrees of motion. It has a consistent performance improvement over all the other algorithms even for sequences with a high degree of motion, although this increase is relatively less when compared to the PSNR increase with the sequences with lower degree of motion. Even when compared to the state-of-the-art MC3DWTr algorithm, the proposed filter is able to outperform it by $\sim 0.25 - 0.4$ dB on average.

Figures 7.3, 7.4 and 7.5 show the PSNR curves obtained using the different algorithm when processing PC, MA, and Ha sequences corrupted to have input PSNRs of 20 dB, 25 dB and 30 dB, respectively. The proposed filter consistently outperforms the spatial as well as wavelet domain filters over the full length of the evaluated sequences. The relative increase observed for the MA and Ha sequences is more than that of the PC sequence on account of the high degree of motion in the PC sequence which causes the TBMAP filter to use only the spatial support for the estimation of the filter parameters.

A visual comparison of the outputs generated by the different filters is presented in Figure 7.6 using one of the frames from the PC video sequence which

	ATM	KNN	RF	MCA	LMMSE	AWA	JKW	MC3DWTr	TBMAP
Input PSNR = 20 dB									
MA	30.37	29.0	29.29	31.21	30.96	29.01	30.65	35.35	34.25
Ha	27.66	27.29	27.31	28.09	27.35	27.75	28.68	29.57	29.75
Fo	27.96	27.75	28.11	28.07	26.14	28.09	27.64	29.85	30.9
PC	22.94	24.14	25.97	26.15	25.5	25.6	24.89	27.99	28.1
Tr	22.38	23.89	24.27	25.53	25.3	25.22	25.59	26.72	26.7
Average	26.25	26.41	26.95	27.81	27.05	27.15	27.48	29.89	29.94
Input PSNR = 25 dB									
MA	33.61	33.45	34.18	34.92	34.57	33.36	32.22	37.32	36.25
Ha	28.74	29.98	30.34	32.04	30.86	31.07	32.37	32.92	34.03
Fo	30.58	30.37	31.66	31.44	30.13	29.85	29.42	32.59	33.52
PC	23.06	25.18	27.93	29.95	28.83	29.08	27.41	31.31	31.75
Tr	22.9	24.8	25.43	29.25	28.74	28.6	28.25	30.08	30.68
Average	27.78	28.75	29.91	31.52	30.63	30.39	29.97	32.84	33.25
Input PSNR = 30 dB									
MA	37.69	37.19	37.53	37.78	37.66	37.3	37.25	38.97	38.61
Ha	29.2	31.38	31.64	35.13	34.66	34.97	35.11	36.05	37.15
Fo	31.25	32.00	33.44	34.65	33.95	31.15	31.72	35.41	36.38
PC	23.06	25.48	28.62	33.2	32.38	32.55	29.85	34.98	36.55
Tr	22.8	25.15	25.8	32.35	32.44	31.97	30.57	33.81	34.21
Average	28.79	30.24	31.41	34.61	34.21	33.59	32.90	35.85	36.58

Table 7.1: PSNR comparison of the proposed algorithm, TBMAP, with several others for the corrupted input test sequences with input PSNRs of 20 dB, 25 dB and 30 dB, resp.

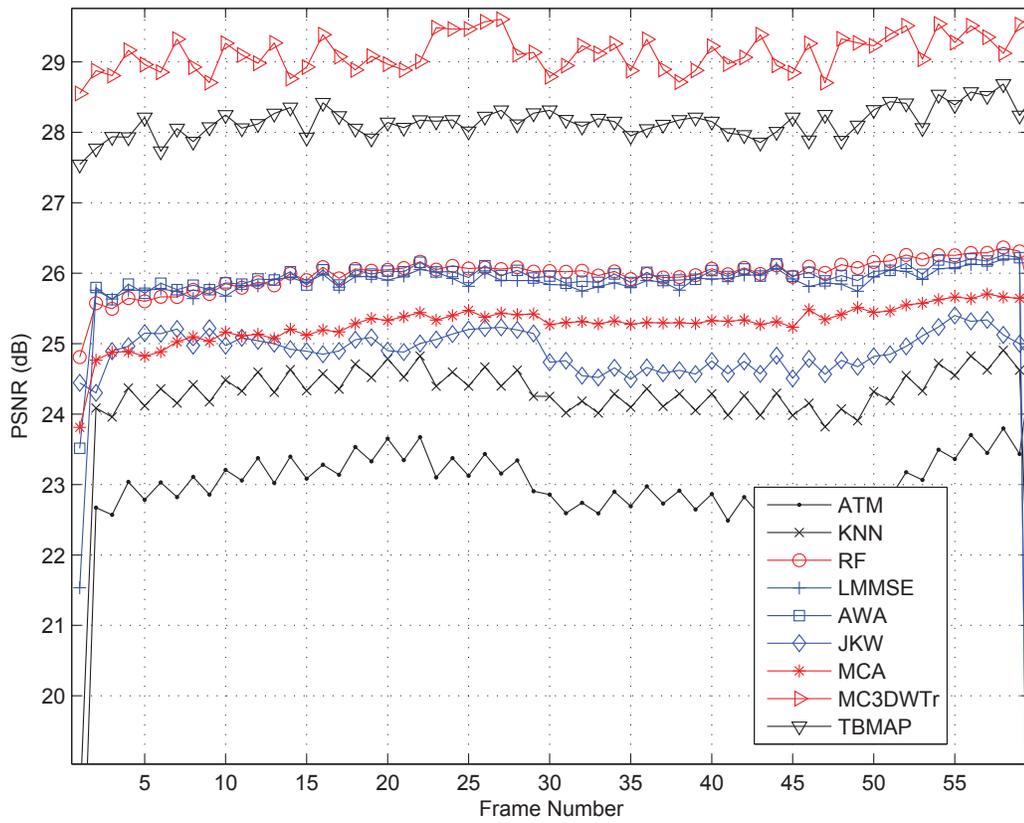


Figure 7.3: PSNR comparison of the different algorithms for *Patrol Car* sequence having an input PSNR of 20 dB

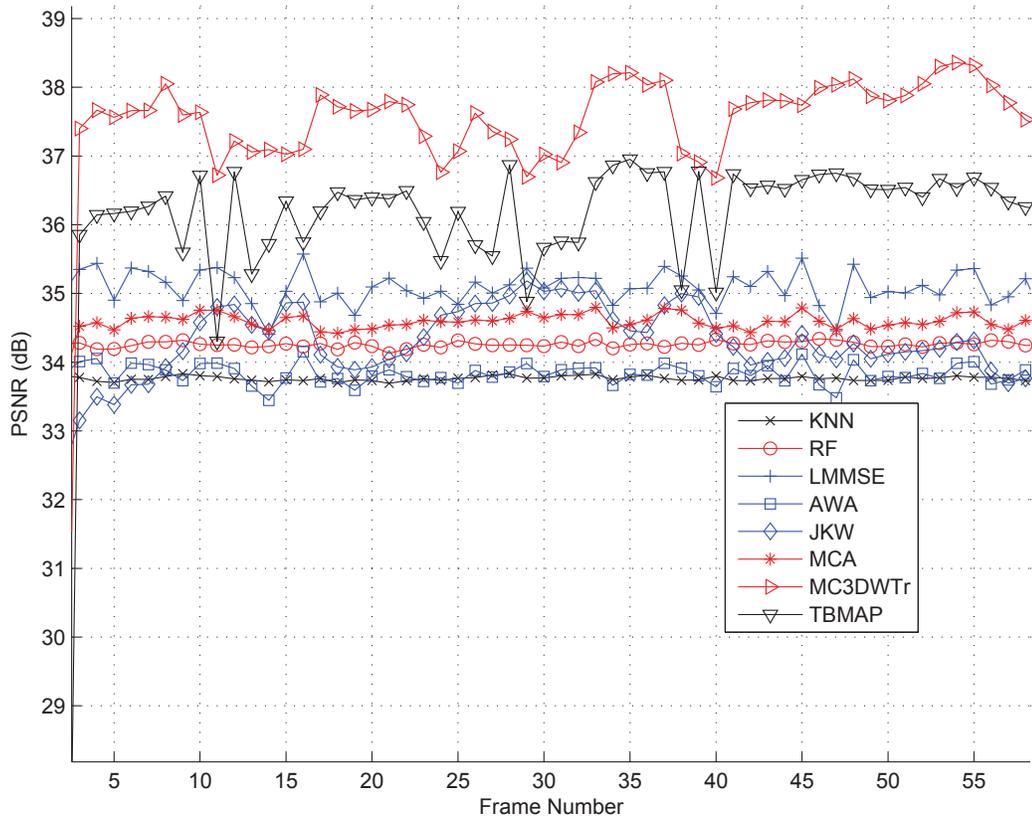


Figure 7.4: PSNR comparison of the different algorithms for *Miss America* sequence having an input PSNR of 25 dB

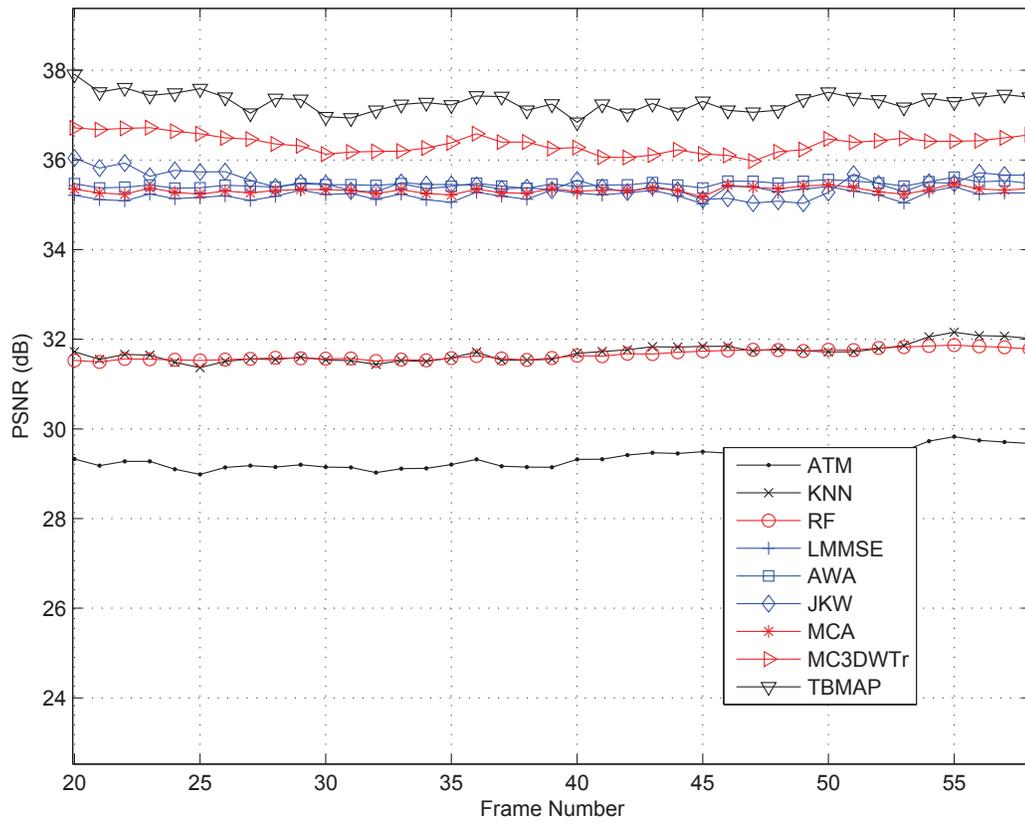


Figure 7.5: PSNR comparison of the different algorithms for *Hall* sequence having an input PSNR of 30 dB

has an input PSNR of 25 dB. As is evident, the proposed filter produces results which are more visually pleasing than filters used for comparison (MC3DWTr, AWA, JKW and MCA). The TBMAP filter is able to reduce noise aggressively due to the bivariate mapping of the subband coefficients and limits the blurring of the detail information in the sequence.

7.5 Conclusion

A spatio-temporally adaptive filter which operates completely in the wavelet domain has been proposed in this chapter. The inter-subband correlations that exist between the successive subbands corresponding to same frame allow the modeling of the parent and child coefficients as a non-Gaussian bivariate distribution. Based on this distribution, a MAP estimator that estimates the noise-free subband coefficients based on the available noisy coefficient and its corresponding parent has been discussed. An adaptive spatio-temporal support has been used for the estimation of the parameters used for the application of the resultant shrinkage function.

Experimental results and simulations have shown the superiority of the proposed filter in video noise reduction over several other algorithms based on both PSNR and visual inspection.

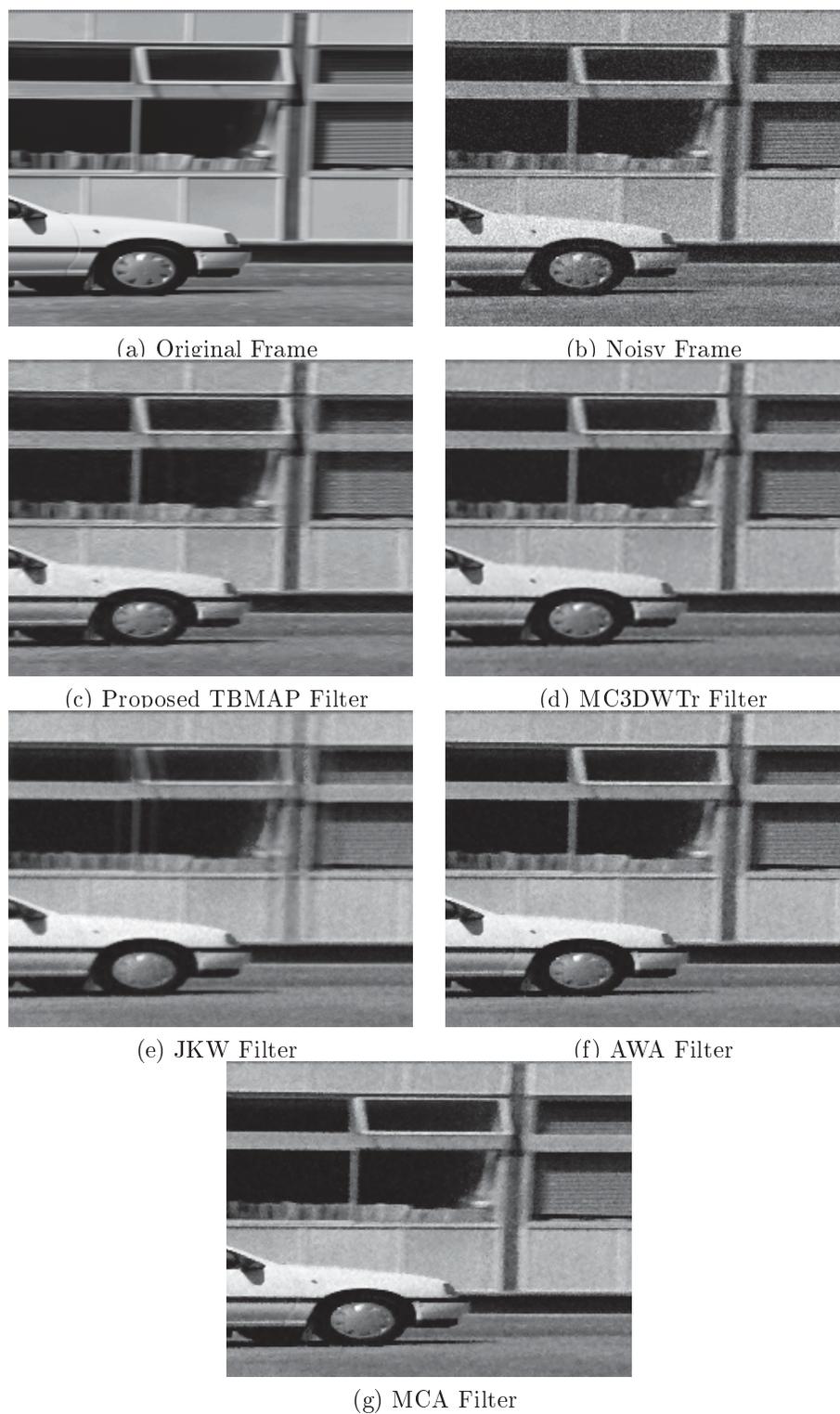


Figure 7.6: Filtered (and zoomed) outputs of one frame of the *Patrol Car* sequence for the different filters

Chapter 8

Conclusion

8.1 Concluding Remarks

The corruption of video sequences with noise is a common phenomenon which is always present to some extent owing to limitations on the hardware (such as sensors), ambient lighting, storage media, etc. With the quick proliferation of the digital videos it has become imperative that this noise be removed from these sequences before either viewing these videos or more importantly, before processing these videos for other applications, such as compression, segmentation and tracking.

Wavelets have been extensively used to denoise the still images and their use has been extended to the domain of video noise reduction. While a considerable amount of research and effort has been spent on the modeling of the subband coefficients corresponding to still images, not much work has been dedicated to the field of video denoising. Videos are different from still images in that there exists additional information in the temporal direction which can be exploited for the enhancement of the videos, thus resulting in better noise reduction. However,

most of the work in video denoising has concentrated on using the spatial filters on individual frames and then using the temporal information in an *ad hoc* way.

Thus, one of the primary goals of this dissertation was to develop a unified framework for spatio-temporal modeling of the wavelet domain representation of the video sequences. In Chapter 3, such a framework was proposed utilizing an *a priori* distribution to model not only the spatial subband coefficients but also the temporal relationship between them. The focus of the work was not to develop priors which model the proposed spatio-temporal framework most accurately, but to show that it is possible to create such an *a priori* framework and model the distributions using a distribution which has a good fit to the empirical data. For the purpose of this dissertation, the spatio-temporal modeling has been done using a generalized Gaussian distribution. Extensive simulations have been performed to verify the validity of this modeling. In the subsequent chapters, the proposed spatio-temporal model for the subband coefficients in a variety of different strategies for noise reduction of the video sequences.

The second goal of this dissertation was to propose the possible alternative strategies to exploit the proposed video model. There are several ways in which the spatio-temporal redundancies can be exploited, such as:

- Reducing the temporal correlations between corresponding subbands before performing any spatial noise reduction.
- Using the spatio-temporal information together in a unified framework that performs optimal noise reduction based on both the spatial and temporal data.
- Exploiting the spatial and temporal redundancies separately using separate filters which operate only along the temporal directions and along the spatial

orientation and then combining the outputs of the two filters adaptively.

- Using the inter-subband correlations in the wavelet representation of a single frame in a multi-variate model and utilizing the spatio-temporal information to enhance its performance.

This dissertation has proposed noise reduction techniques which try to address all of the above strategies of utilizing the video subband data in an optimal and coherent way.

In Chapter 4, a novel spatio-temporal filter for video denoising that operates entirely in the wavelet domain was proposed. For effective noise reduction, the spatial and the temporal redundancies that exist in the wavelet domain representation of a video signal were exploited. The discrete cosine transform (DCT) was used to minimize the redundancies in the temporal direction. After the application of the DCT, the coefficients in the different wavelet domain subbands for the original image sequence were modeled using a prior having a generalized Gaussian (GG) distribution. Based on this prior, filtering of the noisy wavelet coefficients in each subband was carried out using a low-complexity wavelet shrinkage method that utilizes the correlation which exists between subsequent resolution levels.

In Chapter 5, based on the proposed video modeling framework, where the subband coefficients in individual frames as well as the wavelet coefficient difference occurring between two consecutive frames can be modeled using the GG distribution, Bayesian operators, based on minimum mean squared error as well as maximum *a posteriori* estimation, were proposed. These operators estimate the noise-free wavelet coefficients in the current frame conditioned on the noisy coefficients in the current frame and the filtered coefficients in the past frame.

Based on the proposed statistical model, in Chapter 6, another novel noise reduction technique, which exploits the spatial and temporal redundancies that

persist in the wavelet domain representation of the video sequence sequentially has been proposed. The separately processed outputs of a Kalman filter and a spatial Bayesian filter are combined using an adaptive weighted averaging scheme.

In Chapter 7, the inter-scale dependencies in the subband representation of each frame were also modeled using a non-Gaussian bivariate distribution. The parameters for this bivariate distribution were estimated adaptively using the local correlations that exist between neighboring coefficients within each subband. Based on this bivariate distribution a shrinkage function was developed using the maximum *a posteriori* rule. To improve the performance of the filter, information from the adjacent frames was also incorporated in the shrinkage function.

Extensive experimental results for all the presented algorithms were presented and these confirm that the proposed schemes outperform several state-of-the-art spatio-temporal filters in terms of peak signal to noise ratio as well as visual quality. For real-time implementation of the proposed algorithms, in scenarios where only a digital signal processor is available with limited memory resources and computational efficiency is more important, the DCWT and TBMAP algorithms would be ideal. For implementation on a reconfigurable hardware, such as an FPGA, the TMMSE and TMAP algorithms are better suited owing to their possible implementation using lookup tables. Where more computational power is available and very high quality of noise reduction is necessary, the JKMMSE filter would be a better fit.

8.2 Scope for Further Work

The research work presented in this dissertation can be extended in several directions. One of the main focus in this research was to show the feasibility of a

unified framework for the modeling of the spatio-temporal subband coefficients. The GG distribution has been used as the prior for modeling the data in this framework. The effect and accuracy of the use of a different prior, which could make the overall model more robust, can be investigated .

The shrinkage operator based on the non-Gaussian bivariate function incorporates only the spatial correlations. It is possible to extend the same function to incorporate not only the spatial subband correlations but also the temporal ones. It would be worthwhile to investigate as to whether it is possible to extend this filter in a way similar to the one proposed in Chapter 5 wherein the conditional estimation of the present coefficient is done based on the current noisy one as well the corresponding past coefficient.

Bibliography

- [1] G. De Haan. *Video Processing for Multimedia Systems*, University Press Eindhoven, Eindhoven, The Netherlands, 2003.
- [2] C. Chui, *Wavelets*, Academic Press, 1992.
- [3] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.
- [4] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.
- [5] Y. Meyer, *Wavelets: Algorithms and Applicaions*, SIAM, 1993.
- [6] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice-Hall, 1995.
- [7] B. Vidakovic, *Statistical Modeling by Wavelets*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 1999.
- [8] A. Graps, “An Introduction to wavelets,” *IEEE Computational Science and Engineering*, vol. 2, no. 2, 1995.
- [9] G. Strang, “Wavelets,” *American Scientist*, vol. 82, pp. 250-255, April 1994.
- [10] N. Gupta, M. N. S. Swamy, and E. I. Plotkin, “Low-complexity, video noise reduction in wavelet domain,” in *Proc. of IEEE International Workshop on Multimedia Signal Processing*, Sienna, Italy, Sept. 2004, pp. 239–242.

- [11] N. Gupta, E. I. Plotkin, and M. N. S. Swamy, “Bayesian algorithm for video noise reduction in the wavelet domain,” in *Proc. of IEEE International Symposium in Circuits and Systems*, Kobe, Japan, May 2005, vol. 5, pp. 4943–4946.
- [12] N. Gupta, M. N. S. Swamy, and E. I. Plotkin, “Video noise reduction in the wavelet domain using temporal decorrelation and adaptive thresholding,” in *Proc. of IEEE International Symposium in Circuits and Systems*, Island of Kos, Greece, May 2006, pp. 4603–4606.
- [13] N. Gupta, E. I. Plotkin, and M. N. S. Swamy, “Temporally-adaptive MAP estimation for video denoising in wavelet domain,” in *Proc. of International Conference on Image Processing*, Atlanta, USA, Oct. 2006, pp. 1449–1452.
- [14] N. Gupta, M.N.S. Swamy, and E.I. Plotkin, “Wavelet domain–based video noise reduction using temporal discrete cosine transform and hierarchically adapted thresholding,” *IEE Proceedings on Vision, Image and Signal Processing*, vol. 1, no. 1, pp. 2-12, March 2007.
- [15] N. Gupta, M. N. S. Swamy, and E. I. Plotkin, “Motion adaptive video denoising in the wavelet domain based on bivariate shrinkage,” in *Proc. of Canadian Conference on Electrical and Computer Engineering*, Vancouver, Canada, April 2007, pp. 441-444.
- [16] N. Gupta, M. N. S. Swamy, and E. I. Plotkin, “Spatio-temporally adaptive video noise reduction in the wavelet domain using *a priori* modeling in a Bayesian framework,” submitted for review to *IEEE Transaction on Circuits and Systems for Video Technology*.
- [17] B. Vidakovic, “Nonlinear wavelet shrinkage with bayes rules and bayes fac-

- tors,” *Journal of the American Statistical Association*, vol. 93, pp. 173-179, 1998.
- [18] A. Cohen and J. Kovacevic, “Wavelets: the mathematical background,” *Proceeding of the IEEE*, vol. 84, pp. 514-522, 1996.
- [19] S. Mallat, “Wavelets for Vision,” *Proceeding of the IEEE*, vol. 84, no. 4, pp. 604-614, April 1996.
- [20] G. Strang, and V. Strela, “Short wavelets and matrix dilation equations,” *IEEE Transactions on Signal Processing*, vol. 43, pp. 108-115, 1995.
- [21] A. Cohen, I. Daubechies, J. Feauveau, “Biorthogonal bases of compactly supported wavelets,” *Comm. Pure Applied Math*, vol. 45, pp. 485-560, 1992.
- [22] G. P. Nason and B. W. Silverman, “The stationary wavelet transform and some statistical applications,” in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, editors, pp. 281-299, Springer-Verlag, New York, 1995.
- [23] V. Zlokolica, W. Phillips, and D. Van de Ville, “Robust non-linear processing for video processing,” in *Proc. of IEEE International Conference on Digital Signal Processing*, (DSP2002), Santorini , Greece, July 2002, vol. 2, pp. 571-574.
- [24] V. Zlokolica, W. Phillips, and D. Van de Ville, “A new nonlinear filter for video processing,” in *Proc. of IEEE Benelux Signal Processing Symposium*, March 2002.
- [25] D. Martinez and J. S. Lim, “Implicit motion compensated noise reduction of motion video scenes,” in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 375-378, 1985.

- [26] F. Cocchia, S. Carrato, and G. Ramponi, "Design and real-time implementation of a 3-D rational filter for edge preserving smoothing," *IEEE Transactions on Consumer Electronics*, vol. 43, no. 4, Nov. 1997.
- [27] G. Ramponi, "The rational filter for image smoothing," *IEEE Signal Processing Letters*, vol. 3, no. 3, pp. 63-65, March 1996.
- [28] K. Jostschulte and A. Amer, "A new cascaded spatio-temporal noise reduction scheme for interlaced video," in *Proc. of International Conference on Image Processing*, Chicago, USA, Oct. 1998, vol. 2, pp. 493-497.
- [29] V. Zlokolica, A. Pizurica, and W. Philips, "Video denoising using multiple class averaging with multiresolution," *International Workshop VLBV03*, Madrid, Spain, 2003.
- [30] R. D. Turney, A. M. Reza, and J. G. R. Delva, "FPGA implementation of adaptive temporal Kalman filter for real-time video filtering," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, Arizona, USA, March 1999.
- [31] T. S. Huang and Y. P. Hsu, "Image sequence enhancement," in *Image Sequence Analysis*, T. S. Huang (ed.), Chapter 4, pp. 289-309, Springer-Verlag, Berlin, 1981.
- [32] M. I. Sezan, M. K. Ozkan, and S. V. Fogel, "Temporally adaptive filtering of noisy image sequences using a robust motion estimation algorithm," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, May 1991, pp. 2429-2432.
- [33] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald, and P. J. W. Rayner, "Inter-

- polation of missing data in image sequences,” *IEEE Transactions on Image Processing*, vol. 4, no. 11, pp. 1509-1519, Nov. 1995.
- [34] R. Samy, “An adaptive image sequence filtering scheme based on motion detection,” in *Proc. SPIE*, vol. 596, pp. 135-144, 1985.
- [35] M. K. Ozkan, I. Sezan, and A. M. Tekalp, “Adaptive motion compensated filtering of noisy image sequences,” *IEEE Transaction on Circuits and Systems for Video Technology*, Aug. 1993.
- [36] S.V. Fogel, “Estimation of velocity vector field from time-varying image sequences,” *Computer Vision, Graphics: Image Understanding*, vol. 53, pp. 253-287, May 1991.
- [37] M. Bierling, “Displacement estimation by hierarchical block matching,” in *Proc. SPIE Visual Communications and Image Processing*, vol. 1001, pp. 942-951, Nov. 1998.
- [38] J. W. Woods and J. Kim, “Motion compensated spatio-temporal Kalman filter,” in *Motion Analysis and Image Sequence Proc.*, M. I. Sezan and R. L. Lagendijk (eds.), Chapter 12, pp. 349-373, Kluwer Academic Publishers, Boston, March 1993.
- [39] R. Dugad and N. Ahuja, “Video denoising by combining Kalman and Wiener estimates,” in *Proc. IEEE International Conference on Image Processing*, Oct. 1999, vol. 4, pp. 156-159.
- [40] C. Zhu, X. Lin, L.-P. Chau, “Hexagon-based search pattern for fast block motion estimation,” *IEEE Transactions. on Circuits and Systems for Video Technology*, vol. 12, no. 5, pp. 349-355, May 2002.

- [41] C. Zhu, X. Lin, L.-P. Chau, "An enhanced hexagonal search algorithm for block motion estimation," in *Proc. IEEE International Symposium on Circuits and Systems*, May 2003, vol. 2, pp. 392-395.
- [42] S. Zhu, K.K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Transactions on Image Processing*, vol. 9, pp. 287-290, Feb. 2000.
- [43] C.-H. Cheung, L.-M. Po, "A novel rood-diamond search algorithm for fast block motion estimation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2002, vol. 4, pp. 3397-3400.
- [44] K.K Ma, G. Qiu, "An improved adaptive rood pattern search for fast block-matching motion estimation in JVT," in *Proc. IEEE International Symposium on Circuits and Systems*, May 2003, vol. 2, pp. 708-711.
- [45] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425-455, 1994.
- [46] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, pp. 613-627, May 1995.
- [47] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200-1224, December 1995.
- [48] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE International Conference on Image Processing*, Sept. 1996, vol. 1, pp. 379-382.

- [49] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Signal Processing*, vol. 9, no. 9, pp. 1532-1546, Sept. 2000.
- [50] M. Hansen and B. Yu, "Wavelet thresholding via MDL for natural images," *IEEE Transactions on Information Theory (Special Issue on Information Theoretic Imaging)*, vol. 46, pp. 1778-1788, 2000.
- [51] A. Pizurica, *Image Denoising Using Wavelets and Spatial Context Modeling*, Ph.D. Thesis, Ghent University, 2002.
- [52] A. Pizurica, V. Zlokolica, and W. Philips, "Combined wavelet domain and temporal video denoising," in *Proc. IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Miami, FL, July 2003, pp. 334-341.
- [53] P. M. B. V. Roosmalen, S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Noise reduction for image sequences using an oriented pyramid thresholding technique," in *Proc. IEEE International Conference on Image Processing*, Lausanne, Switzerland, 1996, vol. 1, pp. 375-378.
- [54] I. W. Selesnick and K. Y. Li, "Video denoising using 2-D and 3-D dual-tree complex wavelet transforms," in *Proc. SPIE, Wavelets: Appl. Signal Image Processing X*, San Diego, 2003, vol. 5207, pp. 607-618.
- [55] N. Rajpoot, Z. Yao, and R. Wilson, "Adaptive wavelet restoration of noisy video sequences," in *Proc. IEEE International Conference on Image Processing*, Singapore, 2004, vol. 2, pp. 24-27.
- [56] F. Jin, P. Fieguth, and L. Winger, "Wavelet video denoising with regular-

- ized multiresolution motion estimation,” *EURASIP Journal of Applied Signal Processing*, vol. 2006, no. 72705, pp. 1-11, 2006.
- [57] K. Dabov, A. Foi, and K. Egiazarian, “Video denoising by sparse 3-D transform-domain collaborative filtering,” in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Poznan, Poland, pp. 145–149, Sep. 2007.
- [58] S. Yu, M. O. Ahmad, and M. N. S. Swamy, “Video denoising using motion compensated 3-D wavelet transform with integrated recursive temporal filtering,” *IEEE Transactions on Circuits and Systems on Video Technology*, vol. 20, no. 6, pp. 780-791, Jun. 2010.
- [59] E. J. Balster, Y. F. Zheng, and R. L. Ewing, “Combined spatial and temporal domain wavelet shrinkage algorithm for video denoising,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 2, pp. 220-230, Feb. 2006.
- [60] E. J. Balster, Y. F. Zheng, and R. L. Ewing, “Fast, feature-based wavelet shrinkage algorithm for image denoising,” in *Proc. IEEE International Conference on Integration of Knowledge Intensive Multi-Agent Systems*, 2003, vol. 1, pp. 722-728.
- [61] M. K. Mihçcak, I. Kozintsev, K. Ramchandran, and P. Moulin, “Low-complexity image denoising based on statistical modeling of wavelet coefficients,” *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300-303, Dec. 1999.
- [62] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, “Video denoising based on inter-frame statistical modeling of wavelet coefficients,” *IEEE*

- Transactions. on Circuits and Systems for Video Technology*, vol. 17, no. 2, pp. 187-198, Feb. 2007.
- [63] J. Huang, "Statistics of Natural Images and Models," Ph.D. dissertation, Brown Univ., Providence, RI, 2000.
- [64] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vis.*, vol. 18, pp. 17-33, 2003.
- [65] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, Jul. 1989.
- [66] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1773-1784, Aug. 2003.
- [67] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2744-2756, Nov. 2002.
- [68] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886-902, Apr. 1998.
- [69] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning of hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243-250, June 1996.

- [70] J. M. Shapiro, "Embedded image coding using zero-trees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445-3462, Dec. 1993.
- [71] Y. Yoo, A. Ortega, and B. Yu, "Image subband coding using context based classification and adaptive quantization," *IEEE Trans. Image Processing*, vol. 8, pp. 1702-1715, Dec. 1999.
- [72] D. Wei, J. E. Odegard, H. Guo, M. Lang, and C. S. Burrus, "Simultaneous noise reduction and SAR image data compression using best wavelet packet basis," in *Proc. IEEE Intl. Conf. Image Process.*, Washington D.C., October 23-26, 1995.
- [73] J.E. Odegard, H. Guo, C.S. Burrus and R.G. Baraniuk, "Joint compression and speckle reduction of SAR images using embedded zero-tree models," in *Proc. Workshop Image and Multidim. Sig. Process.*, Belize City, Belize, March 3-6, 1996.
- [74] S. G. Chang, B. Yu, and M. Vetterli, "Bridging compression to wavelet thresholding as a denoising method," in *Proc. Conf. Information Sciences Systems*, Baltimore, MD, Mar. 1997, pp. 568-573.
- [75] J. Liu and P. Moulin, "Complexity-regularized image denoising," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. 370-373, Oct. 1997.
- [76] B. K. Natarajan, "Filtering random noise from deterministic signals via data compression," *IEEE Trans. Signal Processing*, vol. 43, pp. 2595-2605, Nov. 1995.
- [77] J. Rissanen, *Stochastic Complexity in Statistical Enquiry*. Singapore: World Scientific, 1989.

- [78] X. Zong, A. F. Laine, and E. A. Geiser, "Speckle reduction and contrast enhancement of echo-cardiograms via multiscale nonlinear processing," *IEEE Trans. Med. Imaging*, vol. 17, pp. 532-540, Aug. 1998.
- [79] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imaging*, vol. 20, no. 8, pp. 772-783, Aug. 2001.
- [80] P. Tsakalides, P. Reveliotis and C. L. Nikias, "Scalar quantization of heavy-tailed signals," *IEE Vis. Image Signal Process.*, vol. 147, no. 5, Oct. 2000.
- [81] N. Gupta, M. N. S. Swamy and E. I. Plotkin, "Low-complexity, hierarchically-adapted wavelet thresholding for image denoising," in *Proc. IEEE TENCON*, vol. 1, pp. 259-262, Nov. 2004.