The SURE-LET Approach to Image Denoising

Florian Luisier

Thèse N° 4566 (January 2010)

Thèse présentée à la faculté des sciences et techniques de l'ingénieur
pour l'obtention du grade de docteur ès sciences
et acceptée sur proposition du jury

Prof. Christian Depeursinge, *président*
Prof. Michael Unser, *directeur de thèse*
Prof. Thierry Blu, *co-directeur de thèse*
Prof. Jean-Christophe Pesquet, *rapporteur*
Prof. Jalal Fadili, *rapporteur*
Prof. Pierre Vandergheynst, *rapporteur*

École polytechnique fédérale de Lausanne—2010

Cover design by Annette Unser
Printing and binding by Repro-EPFL
Typeset with LaTeX
Copyright © 2010 by Florian Luisier
Available at **http://bigwww.epfl.ch/**

Abstract

Denoising is an essential step prior to any higher-level image-processing tasks such as segmentation or object tracking, because the undesirable corruption by noise is inherent to any physical acquisition device. When the measurements are performed by photosensors, one usually distinguish between two main regimes: in the first scenario, the measured intensities are sufficiently high and the noise is assumed to be *signal-independent*. In the second scenario, only few photons are detected, which leads to a strong *signal-dependent* degradation. When the noise is considered as signal-independent, it is often modeled as an additive independent (typically Gaussian) random variable, whereas, otherwise, the measurements are commonly assumed to follow independent Poisson laws, whose underlying intensities are the unknown noise-free measures.

We first consider the reduction of additive white Gaussian noise (AWGN). Contrary to most existing denoising algorithms, our approach does not require an explicit prior statistical modeling of the unknown data. Our driving principle is the minimization of a purely data-adaptive unbiased estimate of the mean-squared error (MSE) between the processed and the noise-free data. In the AWGN case, such a MSE estimate was first proposed by Stein, and is known as "Stein's unbiased risk estimate" (SURE). We further develop the original SURE theory and propose a general methodology for fast and efficient multidimensional image denoising, which we call the SURE-LET approach. While SURE allows the quantitative monitoring of the denoising quality, the flexibility and the low computational complexity of our approach are ensured by a linear parameterization of the denoising process, expressed as a *linear expansion of thresholds* (LET). We propose several pointwise, multivariate, and multichannel thresholding functions applied to arbitrary (in particular, redundant) linear transformations of the input data, with a special focus on multiscale signal representations.

We then transpose the SURE-LET approach to the estimation of Poisson intensities degraded by AWGN. The signal-dependent specificity of the Poisson statistics leads to the derivation of a new unbiased MSE estimate that we call "Poisson's unbiased risk estimate" (PURE) and requires more adaptive transform-domain thresholding rules. In a general PURE-LET framework, we first devise a fast interscale

thresholding method restricted to the use of the (unnormalized) Haar wavelet transform. We then lift this restriction and show how the PURE-LET strategy can be used to design and optimize a wide class of nonlinear processing applied in an arbitrary (in particular, redundant) transform domain.

We finally apply some of the proposed denoising algorithms to real multidimensional fluorescence microscopy images. Such *in vivo* imaging modality often operates under low-illumination conditions and short exposure time; consequently, the random fluctuations of the measured fluorophore radiations are well described by a Poisson process degraded (or not) by AWGN. We validate experimentally this statistical measurement model, and we assess the performance of the PURE-LET algorithms in comparison with some state-of-the-art denoising methods. Our solution turns out to be very competitive both qualitatively and computationally, allowing for a fast and efficient denoising of the huge volumes of data that are nowadays routinely produced in biomedical imaging.

Keywords: image, video, denoising, thresholding, wavelets, multiscale, Stein's unbiased risk estimate (SURE), linear expansion of thresholds (LET), risk, MSE, estimation, approximation, Gaussian, Poisson, microscopy, fluorescence, multidimensional, nonlinear, fast

Résumé

En traitement d'images, le débruitage est une étape essentielle préalablement à n'importe quelles tâches de plus haut niveau telles que la segmentation ou le suivi d'objets, car les dégradations dues au bruit sont inhérentes à tout système d'acquisition physique. Lorsque les mesures sont accomplies par des capteurs photo-sensibles, on distingue principalement deux régimes : dans le premier scénario, les intensités mesurées sont suffisamment élevées et le bruit est considéré comme *indépendant* du signal. Dans le second scénario, seuls quelques photons sont détectés, ce qui conduit à un bruit fortement *dépendant* du signal. Lorsque le bruit ne dépend pas du signal, il est souvent modélisé par une variable aléatoire (typiquement gaussienne) additive et indépendante du signal d'origine, alors que dans le cas contraire, il est communément supposé que les mesures suivent des lois de Poisson indépendantes, dont les données (inconnues) non-bruitées constituent les intensités sous-jacentes.

Nous considérons tout d'abord la réduction du bruit additif, blanc et gaussien. Contrairement à la plupart des algorithmes de débruitage existants, notre approche ne nécessite pas explicitement de modélisation statistique *a priori* du signal inconnu. Notre principe directeur est la minimisation d'un estimateur non-biaisé de l'erreur aux moindres carrés (abrégé MSE en anglais) entre les données traitées et les données non-bruitées ; cet estimateur est exclusivement dérivé à partir des données (observées) bruitées. Dans le cas d'un bruit additif, blanc et gaussien, une telle estimation du MSE fut initialement proposée par Stein et connue sous le nom de "estimation de risque non-biaisée de Stein" (abrégé SURE en anglais). Nous développons davantage la théorie originale du SURE et proposons une méthodologie générale pour le débruitage rapide et efficace d'images multi-dimensionnelles : l'approche SURE-LET. Tandis que le SURE permet un contrôle quantitatif de la qualité du débruitage, la flexibilité et la faible complexité computationnelle de notre approche sont garanties par un paramétrage linéaire du processus de débruitage, exprimé comme une combinaison linéaire de fonctions de seuillage (abrégé LET en anglais). Nous proposons plusieurs fonctions de seuillage (ponctuelles, multi-variées ou multi-canaux) applicables à des transformations linéaires arbitraires (en particulier, redondantes) des données bruitées ; les représentations multi-échelles retiennent principalement notre attention.

Nous transposons ensuite l'approche SURE-LET à l'estimation des intensités d'un processus poissonnien corrompu par du bruit additif, blanc et gaussien. La spécificité de "signal-dépendance" des statistiques poissonniennes requiert la dérivation d'un nouvel estimateur non-biaisé du MSE que nous avons appelé "estimation de risque non-biaisée de Poisson" (abrégé PURE en anglais) et la mise au point de fonctions de seuillage plus adaptatives. Dans un cadre PURE-LET général, nous développons tout d'abord un algorithme rapide de seuillage inter-échelle, restreint à l'utilisation de la transformation en ondelettes de Haar (non-normalisées). Nous levons ensuite cette restriction et montrons comment la stratégie PURE-LET peut être utilisée pour le design et l'optimisation d'une large classe de traitements non-linéaires opérant dans le domaine de transformations linéaires arbitraires (en particulier, redondantes).

Nous appliquons finalement quelques-uns des algorithmes de débruitage proposés sur des images multi-dimensionnelles de microscopie à fluorescence. Cette modalité d'imagerie *in vivo* opère souvent dans des conditions de faible illumination et de brefs temps d'exposition ; en conséquence, les fluctuations aléatoires des mesures des radiations émises par les fluorophores sont fidèlement décrites par un processus de Poisson dégradé (ou non) par un bruit additif, blanc et gaussien. Après une validation expérimentale du modèle statistique de mesure, nous évaluons la performance des algorithmes PURE-LET par rapport aux meilleures méthodes de débruitage disponibles. La solution proposée s'avère très compétitive, tant qualitativement que du point de vue de sa complexité computationnelle, garantissant un débruitage rapide et efficace des énormes volumes de données qui, de nos jours, sont couramment produits en imagerie biomédicale.

Mots-clés : image, video, débruitage, seuillage, ondelettes, multi-échelles, estimation du risque non-biaisée de Stein (SURE), combinaison linéaire de fonctions de seuillage (LET), risque, MSE, estimation, approximation, gaussien, poissonnien, microscopie, fluorescence, multi-dimensionnel, non-linéaire, rapide

A mes parents, Anne & René

Contents

Abstract **i**

Résumé **iii**

List of Notations **xi**

1 Introduction **1**
1.1 Denoising: A Key Step in Image Analysis 1
1.2 Evaluating the Denoising Quality 2
1.2.1 Objective Quality Assessment 3
1.2.2 Subjective Quality Assessment 4
1.3 Survey of Denoising Approaches 5
1.3.1 Filtering 5
1.3.2 Patch-based Approaches 7
1.3.3 PDE and Variational Approaches 9
1.3.4 Transform-Domain Approaches 12
1.4 Contributions 16
1.5 Organization 17

I AWGN Reduction **19**

2 SURE-LET: a Toolkit for AWGN Reduction **21**
2.1 Motivations 21
2.2 Related Work 21
2.3 Stein's Unbiased Risk Estimate (SURE) 23
2.4 Reliability of the SURE Estimate 27
2.5 Optimal Estimator 29
2.6 Linear Expansion of Thresholds (LET) 32
2.7 Summary 33

3 SURE-LET in Orthonormal Wavelet Representation 35
3.1 Motivations . . . 35
3.2 Pointwise Estimator . . . 36
3.2.1 Pointwise SURE . . . 36
3.2.2 Pointwise SURE-LET . . . 37
3.3 Interscale Estimator . . . 40
3.3.1 Building the Interscale Predictor . . . 41
3.3.2 Integrating the Interscale Predictor . . . 45
3.4 Multichannel/Multivariate Estimator . . . 46
3.4.1 Multichannel SURE . . . 50
3.4.2 Multichannel SURE-LET . . . 51
3.4.3 Multichannel Interscale Thresholding . . . 53
3.4.4 From Multichannel to Multivariate Interscale Thresholding . 54
3.5 SURE-LET Estimators: a Comparison . . . 57
3.6 Noise Variance Estimation . . . 57
3.7 Experiments . . . 60
3.7.1 Grayscale Images . . . 60
3.7.2 Color Images . . . 68
3.7.3 Multichannel Images . . . 73
3.8 Summary . . . 76

4 SURE-LET in Non-Orthogonal/Redundant Representation 79
4.1 Motivations . . . 79
4.2 SURE-LET for Transform-Domain Denoising . . . 80
4.3 Pointwise Estimator . . . 81
4.3.1 Case of Non-redundant Transforms . . . 82
4.3.2 Case of Undecimated Filterbank Transforms . . . 83
4.3.3 Pointwise LET . . . 85
4.3.4 Numerical Computation of the Divergence Term $\boldsymbol{\alpha}$ . . . 86
4.3.5 Summary of the algorithm . . . 88
4.3.6 Computational Complexity . . . 89
4.4 Multivariate Estimator . . . 90
4.4.1 Multivariate SURE . . . 90
4.4.2 Multivariate LET . . . 91
4.4.3 Computational Complexity . . . 92
4.5 Multichannel Estimator . . . 92
4.5.1 Multichannel SURE . . . 93
4.5.2 Multichannel LET . . . 93
4.5.3 Multichannel Thresholding . . . 95
4.6 Experiments . . . 95
4.6.1 Transform-Domain vs Image-Domain Optimization . . . 95
4.6.2 Periodic vs Symmetric Boundary Extensions . . . 96
4.6.3 Orthogonal vs Redundant Representation . . . 96

4.6.4 UWT vs Overcomplete BDCT . . . 97
4.6.5 Comparisons with Multiscale Image Denoising Methods . . . 99
4.6.6 Comparisons with Other State-of-the-art Image Denoising Methods . . . 101
4.6.7 Comparisons with State-of-the-Art Color Image Denoising Methods . . . 106
4.7 Summary . . . 107

5 SURE-LET for Fast Video Denoising 111
5.1 Motivations . . . 111
5.2 Multiframe SURE-LET . . . 112
5.2.1 Multiframe SURE . . . 112
5.2.2 Multiframe LET . . . 113
5.3 A Fast Video Denoising Algorithm . . . 114
5.3.1 Global Motion Compensation . . . 114
5.3.2 Local Motion Compensation . . . 114
5.3.3 Multiframe Interscale Wavelet Thresholding . . . 118
5.3.4 Computational Complexity . . . 120
5.4 Experiments . . . 121
5.5 Summary . . . 124

II Poisson Intensities Estimation in AWGN 127

6 PURE-LET: a Toolkit for Estimating Poisson Intensities in AWGN 129
6.1 Motivations . . . 129
6.2 Related Work . . . 130
6.3 Poisson Unbiased Risk Estimate (PURE) . . . 131
6.4 Reliability of the PURE Estimate . . . 134
6.5 Optimal/PURE-LET Estimator . . . 135
6.6 Summary . . . 137

7 PURE-LET in the Unnormalized Haar Wavelet Representation 139
7.1 Motivations . . . 139
7.2 The Unnormalized Haar WaveletTransform . . . 139
7.3 Haar-Wavelet-Domain PURE . . . 141
7.4 Haar-Wavelet-Domain PUREshrink . . . 143
7.5 Haar-Wavelet-Domain PURE-LET . . . 145
7.5.1 Pointwise Thresholding . . . 145
7.5.2 Interscale Estimator . . . 146
7.5.3 Multichannel Estimator . . . 147
7.6 Experiments . . . 151
7.6.1 Poisson Noise Removal . . . 151

7.6.2 Poisson and Gaussian Noise Removal . . . 152
7.7 Summary . . . 156

8 PURE-LET for Arbitrary Transform-Domain Processing 159
8.1 Motivations . . . 159
8.2 Taylor Series Approximation of PURE . . . 159
8.3 PURE for Transform-Domain Denoising . . . 160
8.4 Pointwise Estimator for Undecimated Filterbank Transforms . . . 162
8.4.1 Pointwise PURE . . . 162
8.4.2 Choice of $\widetilde{\mathbf{D}}$ . . . 164
8.4.3 Computation of Transform-dependent Terms . . . 165
8.4.4 Thresholding Function . . . 166
8.4.5 Reliability . . . 167
8.5 Experiments . . . 169
8.5.1 Poisson Noise Removal . . . 169
8.5.2 Poisson and Gaussian Noise Removal . . . 172
8.6 Summary . . . 173

9 Denoising of Fluorescence Microscopy Data 177
9.1 Motivations . . . 177
9.2 Principle of Fluorescence Microscopy . . . 178
9.3 Noise in Fluorescence Microscopy . . . 178
9.3.1 Photon-Counting Noise . . . 179
9.3.2 Measurement Noise . . . 180
9.3.3 Sample-Dependent Noise . . . 180
9.4 Noise Parameters Estimation . . . 181
9.5 Experimental Results . . . 182
9.5.1 2D Fixed Sample . . . 182
9.5.2 3D Sample . . . 186
9.5.3 2D Video Sequence . . . 189
9.6 Summary . . . 190

10 Epilogue 193
10.1 Summary . . . 193
10.2 Perspectives . . . 194

Bibliography 196

Acknowledgement 213

Curriculum Vitæ 215

List of Notations

Symbols & Conventions

$\mathbf{v}$	Vector of N components v_n
v	Any of the components v_n
$\mathbf{M}$	Matrix of $L \times N$ components $m_{l,n}$
$\mathbf{Id}$	Identity matrix
$\mathbf{F}(\cdot)$	Vector function of N components $f_n(\cdot)$
$f(\cdot)$	Any of the scalar functions $f_n(\cdot)$
$\mathbf{J_F}(\cdot)$	Jacobian matrix of the vector function $\mathbf{F}$
$\mathcal{O}(\cdot)$	Order of computational complexity

Operators

$*$	Convolution
$\bullet$	Hadamard product
$\mathscr{E}\{\cdot\}$	Mathematical expectation
∇	Gradient of a scalar function
$\mathrm{div}\{\cdot\}$	Divergence of a vector function
$\mathrm{trace}\{\cdot\}$	Trace of a square matrix
$\mathrm{diag}\{\cdot\}$	Diagonal of a square matrix

Acronyms

AWGN	Additive White Gaussian Noise
BDCT	Block-Discrete Cosine Tranform
BLS	Bayesian Least-Squares
CCD	Charge-Coupled Device
CNR	Contrast-to-Noise Ratio
CS	Cycle-Spin
DCT	Discrete Cosine Transform
DOG	Derivative of Gaussian
FSP	Full Steerable Pyramid
GAT	Generalized Anscombe Transform

GCV	Generalized Cross-Validation
GFP	Green Fluorescent Protein
GSM	Gaussian Scale Mixture
LET	Linear Expansion of Thresholds
MAD	Median of the Absolute Deviation
MAP	Maximum A Posteriori
MMSE	Minimum Mean-Squared error
MSE	Mean-Squared error
OWT	Orthonormal Wavelet Transform
PDF	Probability Density Function
PMT	Photo-Multiplier Tube
PSNR	Peak Signal-to-Noise Ratio
PURE	Poisson's Unbiased Risk Estimate
RGB	Red-Green-Blue
SNR	Signal-to-Noise Ratio
SSIM	Structural Similarity
SURE	Stein's Unbiased Risk Estimate
TI	Translation-Invariant
TV	Total Variation
UWT	Undecimated Wavelet Transform
VST	Variance-Stabilizing Transform

Chapter 1

Introduction

1.1 Denoising: A Key Step in Image Analysis

In image analysis, the extraction of information can be significantly altered by the presence of random distortions called *noise* (see Figure 1.1). The type and energy of this noise naturally depend on the way the images have been acquired or transmitted. The observed image usually consists in a 2D array of pixels: for gray-level images, there is only one channel of light intensity measurement, whereas multispecral (e.g. color) images can have several channels (e.g. RGB: red, green and blue). In most imaging modalities, the measurements are usually performed by a charge-coupled device (CCD) camera which can be seen as a matrix of captors. The pixel value at a particular location is given by the number of photons received by the corresponding captor for a fixed period of time. Most of the noise arises from the fluctuation of the number of incoming photons, but additional perturbations are generated by the thermal instabilities of the electronic acquisition devices and the analog-to-digital conversion. Although the amount of noise actually depends on the signal intensity, it is often modeled as an additive independent (typically Gaussian) random variable, especially when the magnitude of the measured signal is sufficiently high.

There are two main approaches to deal with these unexpected, but also unavoidable, degradations. These are often combined to get a safer solution. The first one is to develop analysis tools that will be robust with respect to noise; the second one, which will retain our attention in this thesis, is to perform a pre-processing step that will consist in *denoising* the data. The trade-off which needs to be optimized is then to reduce the noise level while preserving the key image features.

The vast majority of existing denoising algorithms is specifically designed for the reduction of additive white Gaussian noise (AWGN) in 1D or 2D monochannel data; this considerably reduces their range of application. In particular, denoising the huge multidimensional datasets that are nowadays routinely produced by standard

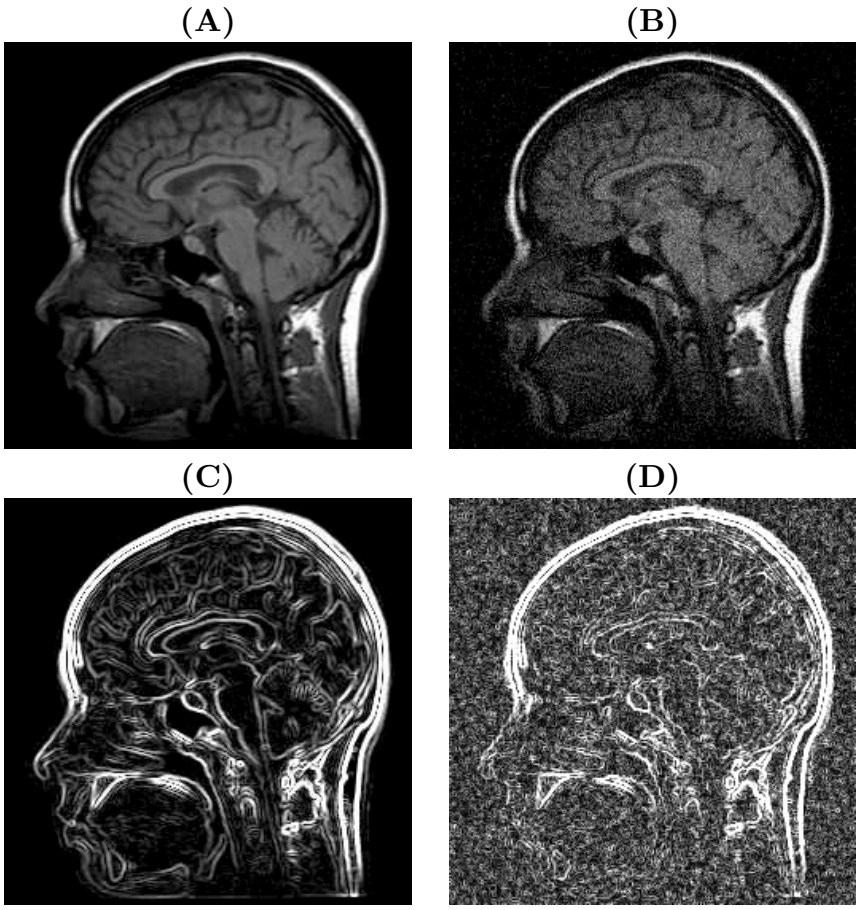


Figure 1.1: **(A)** A noise-free fMRI image. **(B)** A corrupted version of it. **(C)** The result of a classical edge-detector applied on the noise-free image. **(D)** The result of the same edge-detector applied on the noisy image.

biomedical imaging modalities (e.g. microscopy, magnetic resonance imaging (MRI), positron emission tomography (PET),...) remains a challenging task, because it calls for less generic statistical measurement model and computationally more efficient solution.

1.2 Evaluating the Denoising Quality

Evaluating the denoising quality is essential to compare various denoising algorithms or to validate a denoising procedure applied on real data. In this section, we discuss the most popular approaches to image quality assessment.

1.2.1 Objective Quality Assessment

By objective quality assessment, we mean a mathematical measure that quantifies the similarity between the denoised image $\hat{\mathbf{x}} \in \mathbb{R}^N$ and the original noise-free image $\mathbf{x} \in \mathbb{R}^N$. In this respect, the mean-squared error (MSE)

$$\text{MSE} = \frac{1}{N}\|\hat{\mathbf{x}} - \mathbf{x}\|^2 = \frac{1}{N}\sum_{n=1}^{N}(\hat{x}_n - x_n)^2 \tag{1.1}$$

is the most widely used measure of quality. It is often normalized by either the signal energy or the square of the maximum value of the signal and represented in a logarithmic scale yielding either the common signal-to-noise ratio (SNR)

$$\text{SNR} = 10\log_{10}\left(\frac{\|\mathbf{x}\|^2}{\|\hat{\mathbf{x}} - \mathbf{x}\|^2}\right) \tag{1.2}$$

or the popular peak signal-to-noise ratio (PSNR)

$$\text{PSNR} = 10\log_{10}\left(\frac{\|\mathbf{x}\|_\infty^2}{\text{MSE}}\right). \tag{1.3}$$

However, the MSE is considered as a poor visual quality metric [1], mainly due to its non-adaptivity to local signal specificities (intensity, correlation,...). This has led some researchers to design new measures of quality that better correlate with human visual perception [2–6]. The most popular representative of these quality metrics is probably the structural similarity index (SSIM) introduced in [5]. The SSIM between two images $\mathbf{x}$ and $\mathbf{y}$ takes into account three types of similarities:

- the luminance similarity, which involves local measures of the mean of the noisy $\mu_{\mathbf{y}}$ and noise-free images $\mu_{\mathbf{x}}$;
- the contrast similarity, which involves local measures of the variance of the noisy $\sigma_{\mathbf{y}}^2$ and noise-free images $\sigma_{\mathbf{x}}^2$;
- the structural similarity, which also involves local measures of the standard deviation of the noisy and noise-free images, as well as local measures of their correlation $\sigma_{\mathbf{xy}}$.

These similarity measures are combined to yield the following expression of the SSIM:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + C_1)(2\sigma_{\mathbf{xy}} + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + C_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + C_2)} \tag{1.4}$$

where the default values of the two constants C_1 and C_2 are $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$, L being the dynamic range of the pixel values (i.e. $L = 255$ for 8-bit images). Practically, the SSIM is computed inside each $M \times M$ (typically 8×8)

block of the two compared images, yielding a SSIM map. A mean SSIM index can finally be obtained by averaging all the local SSIM values[1].

Perceptual-based quality criterions have two major practical drawbacks: first, none of them is universally adopted by the image processing community; second, optimizing a denoising algorithm based on such criterions is a challenging task, due to their intrinsic structure (non-linear, adaptive and sometimes, parameterized). Yet, we will see throughout this work that, at least in denoising applications, the image exhibiting the lowest MSE is often the one with the highest SSIM score too.

1.2.2 Subjective Quality Assessment

When the original noise-free image is not available (which is the case in real denoising applications), it is not possible to have an objective measure of similarity. However, there are several *reference-free* methods to assess the denoising quality:

- **Human visual assessment:** the simplest way to evaluate the denoising quality is to rely on human judgment. The subjectivity decreases as the the number of human subjects increases. To facilitate the evaluation, their attention is usually drawn to the artifacts visibility and edges sharpness.

- **Method noise** [7]**:** the effect of a denoising algorithm can be evaluated by analyzing the noise guessed by the method, i.e. the difference between the noisy image and the denoised one. Ideally, this residual should have the same statistics as the noise and should not reflect the structure of the original image. While such a requirement is achievable under low noise conditions, it is rarely reached in real-world applications.

- **Contrast-to-noise ratio (CNR):** the contrast-to-noise ratio is a popular measure of quality in biomedical applications [8, 9]. Its use is of less interest for natural images, due to the high diversity of their content. CNR reflects the strength of a feature of interest compared to its environment. Therefore, it gives an indication of the *effective* spatial resolution, i.e. the smallest structure still detectable. It involves the computation of the mean and variance inside a region of interest (ROI) and in the background. Its subjectivity mainly comes from the choice of ROI.

Besides the above mentioned general criterions, other reference-free quality assessment measures have been proposed, often bound to a specific type of distortion [10, 11].

[1] A Matlab code that computes the SSIM map and SSIM index is available at: `http://www.ece.uwaterloo.ca/~z70wang/research/ssim/`.

1.3 Survey of Denoising Approaches

A huge amount of literature is dedicated to image denoising in general, and to additive white Gaussian noise (AWGN) reduction in particular. In this section, we propose a formal classification of the various denoising approaches, based on their underlying working principle. For each class, we further discuss some of its most popular representatives. Note that many other relevant classifications can be envisaged (e.g [7, 12]), due to the hybrid nature of some denoising approaches.

1.3.1 Filtering

Filtering is certainly the simplest way of reducing the noise level in a distorted image. Its denoising efficiency comes from the particular energy distribution of natural images in the frequency domain. Indeed, the noise-free signal spectrum is usually concentrated in low frequencies. It is often assumed to decay like $1/f^{\alpha}$ [13–15], where f is the radial frequency and α is the decay parameter, close to $\alpha = 2$ for most images. In the case of AWGN, the noise power spectrum is constant for all frequencies. Therefore, a mere lowpass filtering of the noisy image can already improve the signal-to-noise ratio (SNR). In this section, we discuss the most standard denoising filters.

Wiener Filter

One of the oldest approach to the image restoration problem is probably the Wiener filter [16]. Suppose that the observed noisy image $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{x} + \mathbf{b}$$

where both the noise-free image $\mathbf{x}$ and the noise $\mathbf{b}$ are independent second-order stationary random processes.

Specifically, the problem is to estimate $\mathbf{x}$ by $\hat{\mathbf{x}} = \mathbf{w} * \mathbf{y}$, and to search for the linear filter $\mathbf{w}$ that minimizes the expected mean-squared error (MSE), i.e.

$$\mathbf{w}_{\text{opt}} = \text{argmin}_{\mathbf{w}} \, \mathscr{E}\left\{\|\mathbf{w} * \mathbf{y} - \mathbf{x}\|^2\right\} \tag{1.5}$$

where $\mathscr{E}\{\cdot\}$ stands for the mathematical expectation.

The solution of this equation is the Wiener filter. Its frequency response is given by:

$$W(\boldsymbol{\omega}) = \frac{S_{\mathbf{x}}(\boldsymbol{\omega})}{S_{\mathbf{x}}(\boldsymbol{\omega}) + S_{\mathbf{b}}(\boldsymbol{\omega})} \tag{1.6}$$

where $S_{\mathbf{x}}(\boldsymbol{\omega})$ is the power spectral density of the noise-free signal $\mathbf{x}$ and $S_{\mathbf{b}}(\boldsymbol{\omega})$ is the power spectral density of the noise $\mathbf{b}$.

The Wiener filter is the minimum mean-squared error (MMSE), space-invariant linear estimator of the signal for stationary images degraded by additive stationary noise. However, when applied to real-world images, it suffers from two main

weaknesses: first, it does not make any distinctions between the treatment of edges and flat-regions due to its space-invariance; second, its performance highly depends on the estimation of the noise-free signal power spectrum. It is *globally* optimum for a Gaussian stationary random signal only. The problem is that natural images do not belong to this particular class of signals. There is a considerable amount of literature that investigates efficient way of estimating the noise-free signal power spectral density in more general cases. The most popular approaches impose some parameterized model (generalized Gaussian or Laplacian, fractal-like, ...) depending on the class of signal to which the noise-free signal is supposed to belong. Due to this difficulty, the Wiener filter is usually outperformed by spatially adaptive and/or non-linear algorithms.

In the simplest *pointwise* space-domain, the equivalent of the Wiener filter amounts to multiplying the centred noisy signal by a constant:

$$\hat{\mathbf{x}} = \Big(1 - \frac{\sigma^2}{\sigma^2_{\mathbf{y}}}\Big)(\mathbf{y} - \mu_{\mathbf{y}}) + \mu_{\mathbf{y}} \tag{1.7}$$

where σ^2, $\sigma^2_{\mathbf{y}}$ and $\mu_{\mathbf{y}}$ are respectively the noise variance, the noisy data variance and the mean of the noisy data. This approach is obviously also transferable to a transformed domain such as wavelet or local discrete cosine transforms (DCTs) [17, 18]. Note that this kind of pointwise linear estimator can be already efficient when applied over regions of interest where the probability density function of the noise-free signal is well approximated by a Gaussian distribution.

A locally adaptive extension of (1.7) has been proposed by Lee in [19]. The idea is to locally compute the mean and variance of the noisy data inside every $M \times M$ overlapping neighborhoods of the image. This adaptive version leads to significant improvements both in terms of MSE and visual quality. This popular algorithm is implemented in the Matlab *wiener2* function.

Other Linear Filters

When there is no information on the noise nor on the noise-free signal statistics, one can still rely on two popular denoising filters: the moving average and the Gaussian smoothing. The level of noise reduction is adjusted via the spatial size of the filter support or the variance of the Gaussian filter. The main drawback of linear filters in general is the blurring of edges.

Non-Linear Filters

Non-linear filters consitute a possible alternative to the standard aforementioned linear filters, the most popular being the median filter. The latter is especially efficient at removing impulsive noise, while giving similar results than linear filters for AWGN removal. Combinations of morphological operators have been also considered

for image denoising [20], in particular when various types of noise are present (e.g. AWGN, salt-and-pepper, speckle,...).

Bilateral Filter

The idea behind the bilateral filtering introduced by Tomasi and Manduchi [21] is to combine *domain* filtering with *range* filtering. Domain filtering is the traditional filtering framework that takes advantage of the similarity of spatially close pixels which are assumed to be more correlated to the centered pixel than more distant pixels. As a consequence, a weighted-averaging of close pixels considerably reduces the noise level. However, in the vicinity of edges, this solution is not satisfying. A better approach for the handling of image discontinuities is adopted by range filtering, which consists in treating together similar pixels, i.e. close in the *range* (intensity) value, without regard to their relative spatial locations. More precisely, Tomasi and Manduchi define range filtering as the *averaging of image values with weights that decay with dissimilarity*. The bilateral filter computes a normalized weighted average of a neighborhood of samples around a central pixel, where the underlying weights are the products of the spatial domain and the range domain weights. The choice of particular decaying functions for the two weights is quite free: the solution proposed by Tomasi and Manduchi is to use Gaussian functions which involve the Euclidian distance for the spatial domain weights and the difference of intensities for the range domain weights. Thus, three parameters control the bilateral filter: the neighborhood size and two parameters setting the decay of the two weight factors. Up to our knowledge, there is no rigorous method for an optimal selection of these parameters.

1.3.2 Patch-based Approaches

The principle of patch-based image denoising is to take advantage of the redundancy of the various structural features present in natural images. Some examples of such redundancy are displayed in Figure 1.2. For instance, an edge can be seen as a succession of similar nearly binary patches. The way of measuring the similarity between image patches and its robustness with respect to noise is at the core of the patch-based approaches.

At a given location $\mathbf{n}$, the estimate $\hat{x}_{\mathbf{n}}$ of the noisy pixel $y_{\mathbf{n}} = x_{\mathbf{n}} + b_{\mathbf{n}}$, is computed as follows:

$$\hat{x}_{\mathbf{n}} = \frac{1}{C} \sum_{\mathbf{k} \in \mathcal{V}(\mathbf{n})} H_h(\mathbf{y_n}, \mathbf{y_k}) G_g(\mathbf{n}, \mathbf{k}) y_{\mathbf{k}} \tag{1.8}$$

where $C = \sum_{\mathbf{k} \in \mathcal{V}(\mathbf{n})} H_h(\mathbf{y_n}, \mathbf{y_k}) G_g(\mathbf{n}, \mathbf{k})$ is a normalizing factor.

$\mathbf{y_n}$ is a vector of pixels (patch) taken from a neighborhood centered around the location $\mathbf{n}$. $\mathbf{k}$ can be the indice of any pixels belonging to the neighborhood $\mathcal{V}(\mathbf{n})$ of

Figure 1.2: Three examples of patches redundancy in the standard *Cameraman* image. Each reference patch is numbered: for clarity, only a few non-overlapping similar patches have been considered here.

the current pixel location $\mathbf{n}$.

The function $H_h(\mathbf{y_n}, \mathbf{y_k})$ measures the photometric similarity between the patches $\mathbf{y_n}$ and $\mathbf{y_k}$. It is usually related to the Euclidian distance between the two patches, as

$$H_h(\mathbf{y_n}, \mathbf{y_k}) = \exp\left(-\frac{\|\mathbf{y_n} - \mathbf{y_k}\|^2}{2h^2}\right)$$

where the parameter h is a threshold parameter, often proportional to the standard deviation of the additive noise.

The function $G_g(\mathbf{n}, \mathbf{k})$ weights the neighboring pixels based on their Euclidian distance to the current pixel location $\mathbf{n}$. It is often a Gaussian kernel with smoothing parameter g.

In the particular case where the patches are reduced to a single pixel, one recovers the *Bilateral Filter* discussed in Section 1.3.1. The *Nonlocal Means Filter* of Buades *et. al* [22, 23] is another famous instance of this class of denoising algorithms.

One of the major challenges in these patch-based approaches is to efficiently fix the various degrees of freedom involved, i.e. the size and shape of the patches (a tradeoff between denoising quality and computational efficiency has to be found), the design of the two weighting functions H_h and H_g, as well as the value of their respective smoothing parameters h and g. Several works along this line have been quite recently undertaken [24, 25].

1.3.3 PDE and Variational Approaches

Anisotropic Diffusion

In the linear formulation of diffusion filtering, the goal is to produce successive version of an original image $I_0(x, y)$ at different resolutions. This is achieved by convolving the original image with Gaussian kernels of increasing variance. It has been shown that this subset of filtered images can be viewed as the solution of the diffusion equation, which can be formulated as:

$$\frac{\partial}{\partial t} I(x, y, t) \quad = \quad \text{div}\left\{c(x, y, t) \nabla I(x, y, t)\right\} \tag{1.9}$$

with the initial condition: $I(x, y, 0) = I_0(x, y)$. $\nabla = \left[\frac{\partial}{\partial x} \; \frac{\partial}{\partial y}\right]^{\mathrm{T}}$ is the spatial gradient operator.

In the simple case where the diffusion conductance $c(x, y, t)$ is a constant C, we talk about linear diffusion with homogenous diffusivity (known as the *heat equation*). This solution leads to a mere Gaussian smoothing of the noisy image. In that case, no distinction is made between the treatment of edges and flat regions of the image.

The solution proposed by Perona and Malik [26] is to use an image-dependent diffusivity $c(x, y, t) = h(|\nabla I(x, y, t)|)$. The diffusion is then controlled by the magnitude of the image gradient at time t:

$$|\nabla I(x, y, t)| = \sqrt{\frac{\partial I(x, y, t)^2}{\partial x} + \frac{\partial I(x, y, t)^2}{\partial y}}$$

Consequently, the edges are better preserved, while flat regions can be heavily smoothed. Their approach is referred to as *non-linear anisotropic diffusion* and is a considerable improvement over conventional Gaussian smoothing. Note that some researchers of the partial differential community call this solution *non-homogenous non-linear isotropic diffusion*, while reserving the terminology of anisotropy for the case where the diffusivity is a diffusion tensor. This approach has been exploited by Weickert in [27] and allows for smoothing along the edges as well.

The delicate aspect of these diffusion-based methods is the choice of an appropriate diffusivity function. Several analytical expressions have been proposed which often involve a conductance parameter. In addition to the type of diffusivity functions and their corresponding conductance, the user has to fix at least two other degrees of freedom: the diffusion speed and the number of iterations. Therefore, the weakest point of the diffusion approach is that it requires the adjustment of several parameters; we are not aware of any rigorous procedure to select the best ones for a given class of images.

Total Variation Minimization

The original total variation (TV) algorithm proposed by Rudin, Osher and Fatemi [28] generated a new trend of research in image denoising. In their approach, denoising is reformulated as a constrained minimization problem. Let $I_0(x,y)$ denote the noisy image observed on the domain Ω of area $|\Omega|$, and described as:

$$I_0(x,y) = I(x,y) + B(x,y) \tag{1.10}$$

where $I(x,y)$ is the original noise-free image and $B(x,y)$ is an additive perturbation independent from $I(x,y)$, such that $\mathscr{E}\{B\} = 0$ and $\mathscr{E}\left\{B^2\right\} = \sigma^2$.

The denoised image is then the solution of

$$\text{argmin}_U \int_\Omega |\nabla U(x,y)|\, \mathrm{d}x\, \mathrm{d}y, \text{ subject to } \frac{1}{|\Omega|}\|U - I_0\|^2 = \sigma^2 \tag{1.11}$$

where

$$\|U - I_0\|^2 \quad = \quad \int_\Omega \left(U(x,y) - I_0(x,y)\right)^2 \, \mathrm{d}x\, \mathrm{d}y$$

Introducing the Lagrange multiplier λ, the problem (1.11) is equivalent to the following (unconstrained) cost functional minimization:

$$J(U) = \int_\Omega |\nabla U(x,y)|\, \mathrm{d}x\, \mathrm{d}y + \frac{\lambda}{2}\int_\Omega \left(U(x,y) - I_0(x,y)\right)^2 \, \mathrm{d}x\, \mathrm{d}y \tag{1.12}$$

with a suitable choice of λ.

∇J can be computed via the Euler-Lagrange equation of (1.12), i.e.

$$\nabla J = \lambda\left(U(x,y) - I_0(x,y)\right) - \text{div}\left\{\frac{\nabla U(x,y)}{|\nabla U(x,y)|}\right\} \tag{1.13}$$

and the minimizer of the functional (1.12) is finally obtained by gradient descent [29]:

$$\frac{\partial}{\partial t}U(x,y,t) = \text{div}\left\{\frac{\nabla U(x,y,t)}{|\nabla U(x,y,t)|}\right\} + \lambda\left(I_0(x,y) - U(x,y,t)\right) \tag{1.14}$$

TV makes a bridge between the pure PDE-based approaches, such as anisotropic diffusion, and the generic class of regularized cost functional discussed in the next section. Indeed, setting $\lambda = 0$ in Equ. (1.14) gives a particular case of anisotropic diffusion, with diffusivity $c(x,y,t) = 1/|\nabla U(x,y,t)|$.

Regularized Cost Functional

The principle here is to find an estimate $\hat{\mathbf{x}} \in \mathbb{R}^N$ of an original signal $\mathbf{x} \in \mathbb{R}^N$ from noisy data $\mathbf{y} \in \mathbb{R}^N$ by minimizing a regularized cost-functional $\mathrm{J}(\mathbf{u},\mathbf{y})$, i.e. $\hat{\mathbf{x}} = \text{argmin}_{\mathbf{u}}\, \mathrm{J}(\mathbf{u},\mathbf{y})$. The cost functional is usually of the following generic form:

$$J(\mathbf{u},\mathbf{y}) = \Psi(\mathbf{u},\mathbf{y}) + \lambda\Phi(\mathbf{u}) \tag{1.15}$$

where Ψ is a data-fidelity term, Φ a regularization term and $\lambda > 0$ a regularization parameter.

The relevant data-fidelity term takes the form:

$$\Psi(\mathbf{u}, \mathbf{y}) = \sum_{n=1}^{N} \psi(u_n - y_n) \tag{1.16}$$

where the function ψ: $\mathbb{R} \to \mathbb{R}$ is typically of the form $\psi(t) = |t|^p$, $p > 0$. The data-fidelity term can also be a more general log-likelihood corresponding to the considered noise model.

The regularization term is commonly of the form:

$$\Phi(\mathbf{u}) = \sum_{n=1}^{N} \varphi([\mathbf{L}_1\mathbf{u}]_n, \ldots, [\mathbf{L}_K\mathbf{u}]_n) \tag{1.17}$$

where $\mathbf{L}_k$ are $N \times N$ matrices, corresponding to discrete versions of some linear operators (e.g. gradient or Laplacian) and $\varphi : \mathbb{R}^K \to \mathbb{R}_+$ is a possibly non-linear differentiable function.

Depending on the choice of particular Ψ and Φ, one recovers a large variety of popular algorithms [30–34]. For instance, the discrete version of the TV algorithm previously introduced can be cast as a representative of this wide class of regularized cost functional. In that case,

- $\mathbf{L}_1$ and $\mathbf{L}_2$ are respectively discrete gradients (finite differences) along the horizontal and vertical directions;
- $\psi(t) = t^2$;
- $\varphi(t_1, t_2) = \sqrt{t_1^2 + t_2^2}$.

When statistical knowledge about the noise and the original noise-free signal are available or assumed, one can adopt a Bayesian formulation which leads to the same general form of cost-functional presented in (1.15). Indeed, maximizing the posterior distribution

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \tag{1.18}$$

amounts to minimizing the following log-likelihod functional:

$$J(\mathbf{u}, \mathbf{y}) = -\log p(\mathbf{y}|\mathbf{u}) - \log p(\mathbf{u}) \tag{1.19}$$

The minimizer of (1.19) is known as the maximum a posteriori (MAP) estimator. It is one of the most widely used solution in transform-domain approaches (see Section 1.3.4). Note that there is also a penalized likelihood framework [35, 36], where the probability term $-\log p(\mathbf{u})$ is replaced by a more general penalty $\lambda\Phi(\mathbf{u})$.

1.3.4 Transform-Domain Approaches

The transform-domain approach constitutes another wide class of denoising algorithms. Initially, the Fourier transform was naturally considered, but the potential of other transforms, such as the DCT early chosen by Yaroslavsky [17], has been soon exploited for image denoising. The advent of powerful multiresolution tools such as the wavelet transform [37–39] has then contributed a lot to the popularity of this denoising strategy. A "good" transform candidate for denoising applications should enjoy (at least some of) the following properties:

- **Invertibility:** Since the denoising output will be usually visualized and/or further analysed, it must be represented in the same domain as the input. Hence, the transform must be perfectly invertible; i.e. the inversion procedure should not add any distortions.

- **Linearity:** A linear transformation is highly preferable for the tractability of the signal and noise dependencies in the transform-domain. Linearity also facilitates the perfect invertibility of the transformation and usually lowers its computational complexity. Non-linearities are preferably reserved for the processing itself.

- **Computational efficiency:** In most applications, denoising is routinely performed prior to higher-level image processing. Therefore, its computational cost must be ideally negligible compared to that of more evolved operations such as segmentation, detection or tracking. For fast denoising, the following scenarios are usually considered: either a low-cost transform precedes a sophisticated processing or a simple (i.e. often pointwise) operation follows a more involved transformation. Combining complicated processing with high-complexity transforms is often prohibitive in real-world applications.

- **Decorrelation ability:** One of the main reasons for processing the noisy data in a transform-domain is to remove inter-pixel dependencies. Hence, for certain class of signals, *pointwise* operations in an appropriate transform-domain can be already very efficient (e.g. Wiener filter for stationary processes). The decorrelation ability of the transform is also crucial for Bayesian denoising approaches: it often allows an easier explicit description of the prior on the noise-free data. However, one should keep in mind that "decorrelation" is *only* statistically synonymous with "independence" in the case of a Gaussian distribution.

- **Energy compaction:** It is highly desirable that the transform concentrates the energy of the significant image features in a small number of high-magnitude transform coefficients. This property is primordial in image compression. In this field, popular transforms are the 8×8 block discrete cosine trans-

form (BDCT) [40] used in JPEG, and the Cohen-Daubechies-Feauveau 9/7 biorthogonal wavelet transform [39], exploited in JPEG2000.

Energy compaction is closely linked to the *sparsity* power of the transform. A linear transform $\mathbf{A} \in \mathbb{R}^{M \times N}$ is said to yield a sparse representation of a vector $\mathbf{x} \in \mathbb{R}^N$, if $\|\mathbf{Ax}\|_0 = \#\{m : [\mathbf{Ax}]_m \neq 0\} \ll M$. However, in practice, a strict sparse representation is very difficult (if not impossible) to achieve for natural images [41]. Therefore, the ℓ_0 norm is usually replaced by a ℓ_p norm, $0 < p \leq 1$. Since $0 < p < 1$ leads to non-convex norms, only the ℓ_1 norm is considered as sparsity promoter in practical denoising applications. If some technical constraints (e.g. related to the *mutual coherence* of the columns of $\mathbf{A}$) are satisfied, then the ℓ_1 norm minimization also leads to the sparsest solution [42].

To achieve a good energy compaction of the input signal, the transform basis functions must efficiently occupy the time-frequency plan (see Figure 1.3). The two extreme representations are the image-domain (perfect spatial localization) and the Fourier-domain (perfect frequency localization). Due to the Heisenberg uncertainity principle, simultaneous perfect time-frequency localization is unachievable. Therefore, the design of any transforms requires a trade-off between accuracy in time and frequency localization. For instance, the basis functions of the BDCT uniformly cover the time-frequency plan, while the wavelet transform offers a multiresolution representation, thanks to a non-uniform tilling of the time-frequency plan.

- **Shift Invariance:** Shift invariance refers to the commutativity of the transform with respect to translation of the input signal. The major drawback of the standard transforms early considered for image denoising (DCT, non-overlapping BDCT, basis of wavelets) is their sensitivity to spatial shifts of the input data. The resulting artifacts (local Gibbs-like phenomenon) are predominant near image discontinuities such as edges. The so-called *blocking* artifacts (visibility of the boundaries between processed blocks) are typical to non-overlapping block-transforms.

  Shift invariance can be achieved by a general procedure introduced by Donoho and Coifman [43] coined *cycle-spinning*. It consists of the following steps:

  1. Shift the input signal.
  2. Apply the transform.
  3. Denoise the shifted transformed coefficients.
  4. Apply the inverse transform.
  5. Shift back the denoised output.
  6. Repeat steps 1 to 5 for a range of shifts and average the various denoised outputs.

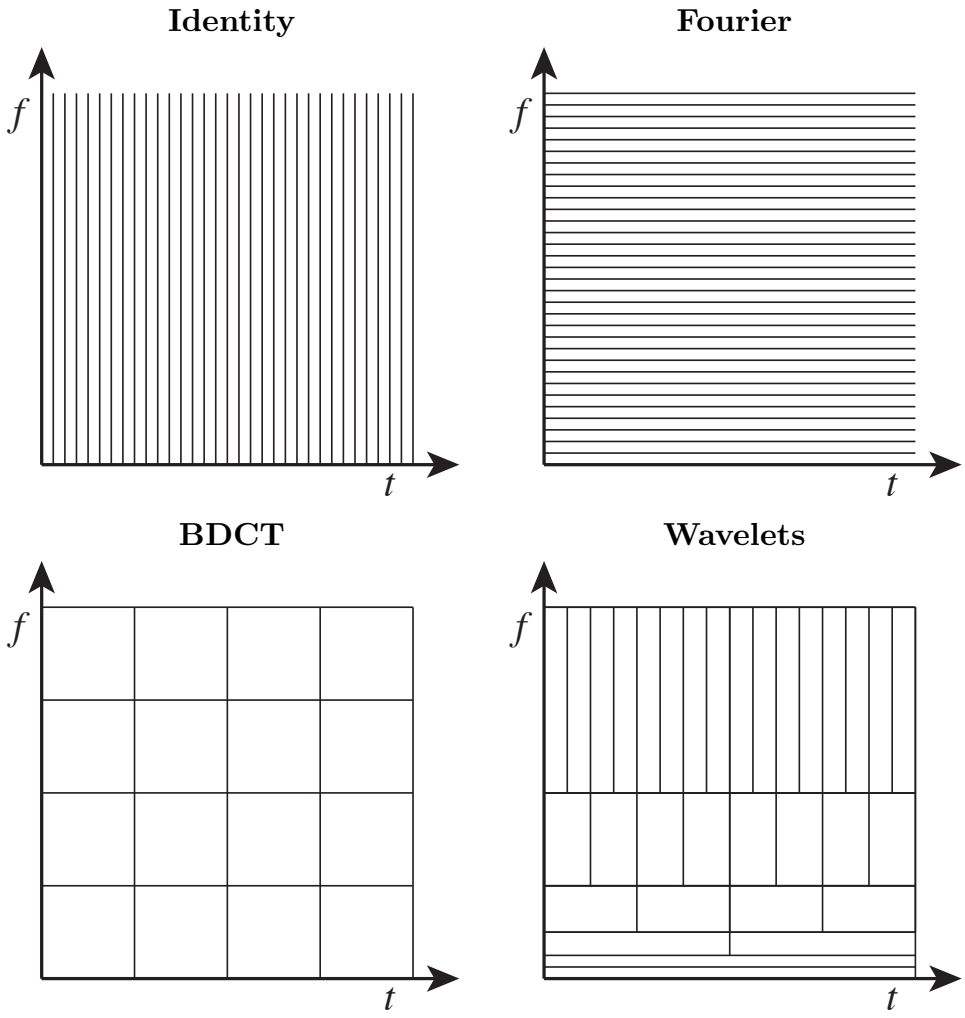


Figure 1.3: Schematic time-frequency occupation of the basis functions of some popular transforms.

Perfect shift-invariance can be achieved if all cyclic shifts are considered. Since this procedure is relatively time-consumming, only a small number of shifts (~ 20) are performed in practice.

There are naturally other transform-specific ways of increasing or achieving shift-invariance: for instance, by considering overlapping blocks for block-transforms (e.g. sliding window DCT [17]). In the case of wavelets, better shift-invariance can be obtained with:

– sharper bandpass filters, the optimal case being the use of ideal filters, i.e. Shannon wavelets;

  - complex or dual-tree wavelets;
  - undecimated wavelet transform.

  We refer the reader to [44] for a more comprehensive study on how to increase the shift-invariance of the discrete wavelet transform.

- **Rotation Invariance:** Textures and edges are usually present in natural images at various orientations. Rotation invariance can be useful in image denoising to better capture and more efficiently represent these oriented features. Unfortunately, rotation invariance is more difficult to achieve than shift-invariance. Indeed, a general procedure similar to cycle-spinning would require interpolation of the input image sampling grid, and thus introduce artificial correlations between neighboring pixels. One way of avoiding interpolation, is to consider rotation along concentric rectangles: this solution has been described in [45] and coined *quasi-circular* rotation.

  In the wavelet community, rotation-invariance is often related to directional sensitivity/selectivity. The standard critically-sampled wavelet transform is sensitive to only two preferential orientations (higher response for horizontal and vertical features), due to the use of a separable basis. Higher directional sensitivity can be achieved by the so-called *shiftable* or *steerable* multiscale transforms [46, 47]. The key ingredients of such transforms are the steerable filters [48–50], which are essentially directional derivatives operators. Since the early work of Simoncelli *et. al.*, numerous directional wavelet transforms have been designed, stemming from either real (e.g. [51], *Curvelets* [52, 53], *Directionlets* [54], *Contourlets* [55]) or complex basis functions (e.g. [56–59]). Instead of increasing the directional selectivity, one can envisage the exact opposite, i.e. considering transforms which do not favor any particular orientations, such as isotropic wavelets [60]. Yet, isotropic transforms are usually outperformed by the directional ones, except for some particular datasets, such as the ones encountered in astronomy [61].

Some of the properties evoked above can be enhanced by considering *data-adaptive* transformations. For instance, the Karhunen-Loève transform can perfectly decorrelate the data. Its empirical equivalent, principal component analysis (PCA), has been therefore exploited in some denoising algorithms (e.g. [62, 63]). Recently, a shape-adaptative BDCT [64] has been efficiently used for image denoising by Foi *et. al.* [65]. Adaptive wavelet transforms (e.g. *Bandelets* [66], *Grouplets* [67]) have been designed to better represent the specific geometry of a given image. Adaptive partitioning of images has been also considered for a sparser representation of edges in [68](*Wedgelets*) and [69](*Platelets*).

1.4 Contributions

The general goal of this thesis is to design denoising algorithms that satisfy the following requirements:

- **Efficiency:** the proposed solutions have to be competitive with the state-of-the-art denoising methods, with respect to an objective measure of quality (e.g. PSNR).
- **Computational cost:** the proposed algorithms should be faster than (or at least as fast as) the most efficient denoising algorithms available. A modest computational memory usage is also highly desirable in order to process large datasets.
- **User intervention:** the proposed algorithms should request a minimal intervention from the user; ideally, the whole denoising procedure should be fully automated.
- **Aimed application:** the main field of application is the denoising of biomedical volumes, with a special focus on multidimensional fluorescence microscopy data.

Along the lines leading to the fulfillment of the above requirements, this thesis brings the following contributions:

- **Theoretical contributions:** To come up with efficient denoising algorithms, we need to achieve optimal performances with respect to an objective measure of quality. Therefore, the driving force behind our approach is the minimization of the mean-squared error (MSE). Since the computation of the MSE requires the knowledge of the unknown noise-free signal, this criterion can not be optimized in real applications. Instead, we propose to minimize a robust unbiased estimate of the MSE. For AWGN, such an estimate has been established in the 1980's by Charles Stein and is nowadays known as Stein's unbiased risk estimate (SURE) [70]. In the denoising community where the performance is almost universally measured in term of MSE, SURE has surprisingly remained under-exploited, the Bayesian approach being adopted by the vast majority of researchers (to the noteworthy exceptions of Donoho *et. al.* in [71], Pesquet *et. al.* in [72–74] and Zhang *et. al.* in [75]). Remarkably, and contrary to Bayesian approaches, the derivation of unbiased estimates of the MSE does not require the statistical modeling of the underlying noise-free signal. The latter can even be considered as deterministic.

  In this work, we propose to revitalize the SURE theory. We derive several expressions of SURE bound to a type of processing (pointwise or multivariate) and a type of transformation (orthogonal or not). We then show that unbiased estimates of the MSE can be obtained beyond the original framework of AWGN

defined by Stein. In particular, we give the expressions of an exact and an approximated unbiased estimate of the MSE for a Poisson process degraded by an AWGN, a quantity that we named *PURE* in reference to Poisson unbiased risk estimate.

- **Algorithmic contributions:** Most of the existing denoising algorithms require the optimization of several non-linear parameters and/or involve sophisticated redundant transforms. Consequently, their computational burden is usually quite heavy. To achieve a high-quality denoising in a low computation time, we propose a generic form of denoising process, expressed as a transform-domain *linear expansion of thresholds* (LET). The optimal linear parameters of this expansion are then the ones that minimize an unbiased estimate of the MSE. Thanks to the quadratic form of such MSE estimates, the whole parameters optimization boils down to the solution of a *linear* system of equations. In this LET framework, fast transform-domain denoising algorithms can then be devised.

- **Practical contributions:** Our practical contributions are twofold. First, by running a comprehensive set of denoising experiments for grayscale/multichannel images/multi-dimensional data, we show that SURE-LET approaches constitute a competitive alternative to the main stream generated by the Bayesian standpoint. Second, by deriving PURE-LET estimators based on a realistic observation model (i.e. related to the physic of the acquisition process), we provide a fast and efficient tool for denoising real fluorescence microscopy volumes.

1.5 Organization

The first part of this thesis is dedicated to additive white Gaussian noise (AWGN) reduction. Although numerous works have been already undertaken in this (yet still active) field of research, we present in Chapter 2 an innovative data-driven procedure (*SURE-LET*) that constitutes a competitive alternative to the state-of-the-art denoising strategies.

In Chapter 3, we show how this general SURE-LET denoising procedure can be efficiently applied to *orthonormal* wavelet thresholding. In particular, we propose an original way of considering and integrating the inter- and intrascale wavelet dependencies. A natural vector/matrix generalization of the SURE-LET strategy is also devised for the denoising of multichannel images.

In Chapter 4, we extend the SURE-LET strategy to arbitrary (included redundant) transform-domain denoising. We point out that, contrary to the particular case of orthogonal transform, the SURE-based parameters optimization has to be performed in the image domain, to ensure a global SURE optimum. Some examples

of redundant transformed-domain thresholding are presented for both mono- and multichannel image denoising.

In Chapter 5, we devise a low-complexity, yet remarkably efficient, SURE-LET algorithm for video denoising. In particular, we propose a simple motion-compensation procedure that can be well-integrated into our SURE-LET framework. Extensive comparisons with the state-of-the-art *redundant* wavelet-based video denoising schemes show that the proposed *non-redundant* solution is already competitive.

In the second part of this thesis, we take up the problem of estimating Poisson intensities degraded by AWGN. Following the same philosophy as in the AWGN case, we derive in Chapter 6 a generic procedure that we named *PURE-LET*.

Chapter 7 is dedicated to the adaptation of the PURE-LET strategy to Haar wavelet-domain thresholding, because this is the only case where a rigorous independent wavelet subband parameters optimization is possible. We show that, although restricted to the non-redundant Haar wavelet representation, the proposed PURE-based denoising algorithms are already competitive with most state-of-the-art approaches, in particular those that combine a variance-stabilizing transform with an AWGN denoiser.

Unfortunately, PURE turns out to be impractical to use in arbitrary transform-domain denoisng. In Chapter 8, we thus give the expression of an approximated PURE, which can be effectively used in practice. We demonstrate that this approximation remains close to the actual MSE. A novel undecimated Haar wavelet thresholding is then devised; it is found to compare favorably to the state-of-the-art techniques for Poisson intensities estimation.

Several experiments on real multidimensional fluorescence microscopy data are presented in Chapter 9. A simple procedure to estimate the various parameters involved in the considered measurement model is also proposed, and validated using multiple acquisitions of the same fixed sample. Our concluding remarks, as well as some perspectives, are finally reported in Chapter 10.

Part I

AWGN Reduction

Chapter 2

SURE-LET: a Toolkit for AWGN Reduction

2.1 Motivations

Given some noisy measurements, the goal of any denoising algorithm is to find the *best* estimate of the underlying noise-free signal. The key point is then to quantify how close to the original signal a given estimate is. We have already discussed in Chapter 1 several measures of quality assessment. From a practical point of view, the mean-squared error (MSE) clearly emerges as the best candidate, due to its appealing mathematical properties (quadratic, symmetric, differentiable, invariant to unitary transforms). In this chapter, we thus propose to consider the minimization of the MSE as the driving principle of our denoising strategy.

2.2 Related Work

The toolkit that will be introduced in this chapter is essentially meant for transform-domain denoising in general, and wavelet-based in particular. In this section, we thus briefly review the most standard, as well as the most efficient, wavelet-domain techniques designed for additive white Gaussian noise (AWGN) reduction.

The most straightforward way of distinguishing information from noise in the wavelet domain consists in thresholding the wavelet coefficients. Of the various thresholding strategies, *soft-thresholding* is the most popular and has been theoretically justified by Donoho and Johnstone [71]. These authors have shown that the shrinkage rule is near-optimal in the minimax sense; they also provided the expression of the optimal threshold value T, called the "universal threshold", as a function of the noise power σ^2 when the number of samples N is large: $T = \sqrt{2\sigma^2 \log N}$.

The use of the universal threshold to denoise images in the wavelet domain is known as *VisuShrink* [76].

Yet, despite its theoretical appeal, minimax is different from MSE as a measure of error. A lot of work has been gone into proposing alternative thresholding strategies that behave better in terms of MSE than *VisuShrink* [56, 77–79]. Donoho and Johnstone themselves acknowledged the poor MSE performance of *VisuShrink* and suggested to choose the optimal threshold value T by minimizing Stein's unbiased risk estimator (SURE) [70] when the data fail to be sparse enough for the minimax theory to be valid. This hybrid approach has been coined *SUREshrink* by their authors [71]. Without challenging the soft-thresholding strategy, alternative threshold value selections have been proposed as well. One of the most popular was introduced by Chang *et al.*, who derived their threshold in a Bayesian framework, assuming a generalized Gaussian distribution for the wavelet coefficients. This solution to the wavelet denoising problem is known as *BayesShrink* [80] and has a better MSE performance than *SUREshrink*.

Beyond the pointwise approach, more recent investigations have shown that substantially larger denoising gains can be obtained by considering the intra- and interscale correlations of the wavelet coefficients. In addition, increasing the redundancy of the wavelet transform is strongly beneficial to denoising performance, a point to which we will come back later in this work. Among the many denoising algorithms to date, we would like to emphasize the following ones:

- *Portilla et al.* [81][1]: The authors' main idea is to model the neighborhoods of coefficients at adjacent positions and scales as a Gaussian scale mixture (GSM); the wavelet estimator is then a Bayesian least squares (BLS). The resulting denoising method, consequently called *BLS-GSM*, is the most efficient multiresolution-based approach in terms of peak signal-to-noise ratio (PSNR).

- *Pižurica et al.* [82][2]: Assuming a generalized Laplacian prior for the noise-free data, the authors' approach called *ProbShrink* is driven by the estimation of the probability that a given coefficient contains significant information (notion of "signal of interest").

- *Sendur et al.* [83, 84][3]: The authors' method, called *BiShrink*, is based on new non-Gaussian bivariate distributions to model interscale dependencies. A nonlinear bivariate shrinkage function using the maximum a posteriori (MAP) estimator is then derived. In a second paper, these authors have extended their approach by taking into account the intrascale variability of wavelet coefficients.

[1] available at `http://www.io.csic.es/PagsPers/JPortilla/denoise/software/index.htm`.
[2] available at `http://telin.ugent.be/~sanja/`.
[3] available at `http://taco.poly.edu/WaveletSoftware/denoise2.html`.

These techniques have been devised for both redundant and non-redundant transforms. Throughout this work, we will compare our results with those obtained by these methods, which are representative of the state-of-the-art in wavelet denoising.

2.3 Stein's Unbiased Risk Estimate (SURE)

We consider here the following standard denoising problem. Suppose we observe a noisy realization $\mathbf{y} = [y_n]_{n=1\dots N}$ of an original signal $\mathbf{x} = [x_n]_{n=1\dots N}$ distorted by an additive white Gaussian noise (AWGN) $\mathbf{b} = [b_n]_{n=1\dots N}$ of variance σ^2, i.e.

$$\mathbf{y} = \mathbf{x} + \mathbf{b}, \text{ with } \mathbf{b} \sim \mathcal{N}(0, \sigma^2 \mathbf{Id}) \tag{2.1}$$

Our goal is to find an estimate $\hat{\mathbf{x}} = \mathbf{F}(\mathbf{y}) = [f_n(\mathbf{y})]_{n=1,\dots N}$ that minimizes the mean-squared error (MSE) defined by:

$$\text{MSE} = \frac{1}{N}\|\hat{\mathbf{x}} - \mathbf{x}\|^2 = \frac{1}{N}\sum_{n=1}^{N}(\hat{x}_n - x_n)^2 \tag{2.2}$$

Since we do not have access to the original signal $\mathbf{x}$, we cannot compute the above *Oracle* MSE. However, without any assumptions on the noise-free data, we will see that it is possible to replace this quantity by an unbiased estimate which is a function of $\mathbf{y}$ only. This has an important consequence: contrary to what is frequently done in the denoising literature (Bayesian approaches), the noise-free signal is not modeled as a random process in our framework (we do not even require $\mathbf{x}$ to belong to a specific class of signals). Thus, the observed randomness of the noisy data only originates from the AWGN $\mathbf{b}$.

The following lemma, which states a version of Stein's lemma [70], shows how it is possible to replace an expression that contains the unknown data $\mathbf{x}$ by another one with the same expectation, but involving the known data $\mathbf{y}$ only:

Lemma 1. *Let $\mathbf{F}(\mathbf{y}) = [f_n(\mathbf{y})]_{1\leq n\leq N}$ be an N-dimensional vector function such that, for $n = 1 \dots N$,*

- *$f_n(\mathbf{y})$ is (weakly) differentiable w.r.t. y_n;*
- *$f_n(\mathbf{y})$ is bounded by some* fast increasing function*, typically such that: $|f_n(\mathbf{y})| \leq \text{cst} \cdot \exp(\|\mathbf{y}\|^2/2\alpha^2)$, where $\alpha > \sigma$.*

Then, under the AWGN assumption, the expressions $\mathbf{F}(\mathbf{y})^{\mathrm{T}}\mathbf{x}$ and $\mathbf{F}(\mathbf{y})^{\mathrm{T}}\mathbf{y} - \sigma^2 \text{div}\{\mathbf{F}(\mathbf{y})\}$ have the same expectation:

$$\mathscr{E}\left\{\sum_{n=1}^{N} f_n(\mathbf{y}) x_n\right\} = \mathscr{E}\left\{\sum_{n=1}^{N} f_n(\mathbf{y}) y_n\right\} - \sigma^2 \mathscr{E}\Bigg\{\underbrace{\sum_{n=1}^{N} \frac{\partial f_n(\mathbf{y})}{\partial y_n}}_{\text{div}\{\mathbf{F}(\mathbf{y})\}}\Bigg\} \tag{2.3}$$

where $\mathscr{E}\{\cdot\}$ stands for the mathematical expectation operator.

Proof. We use the fact that a white Gaussian probability density $q(b_n)$ satisfies $b_n q(b_n) = -\sigma^2 q'(b_n)$. Thus, denoting by $\mathscr{E}_{b_n}\{\cdot\}$ the partial expectation over the n^{th} component of the noise, we have the following sequence of equalities:

$$\begin{aligned}
\mathscr{E}_{b_n}\{f_n(\mathbf{y})x_n\} &= \mathscr{E}_{b_n}\{f_n(\mathbf{y})y_n\} - \mathscr{E}_{b_n}\{f_n(\mathbf{y})b_n\} \\
&= \mathscr{E}_{b_n}\{f_n(\mathbf{y})y_n\} - \int_{\mathbb{R}} f_n(\mathbf{y}) b_n q(b_n) \mathrm{d}b_n \\
&= \mathscr{E}_{b_n}\{f_n(\mathbf{y})y_n\} + \sigma^2 \int_{\mathbb{R}} f_n(\mathbf{y}) q'(b_n) \mathrm{d}b_n \\
&= \mathscr{E}_{b_n}\{f_n(\mathbf{y})y_n\} - \sigma^2 \int_{\mathbb{R}} \frac{\partial f_n(\mathbf{y})}{\partial y_n} q(b_n) \mathrm{d}b_n \quad \text{(by parts}^4\text{)} \\
&= \mathscr{E}_{b_n}\{f_n(\mathbf{y})y_n\} - \sigma^2 \mathscr{E}_{b_n}\left\{\frac{\partial f_n(\mathbf{y})}{\partial y_n}\right\}
\end{aligned}$$

Now, taking the expectation over the remaining components of the noise, we get

$$\mathscr{E}\{f_n(\mathbf{y})x_n\} = \mathscr{E}\{f_n(\mathbf{y})y_n\} - \sigma^2 \mathscr{E}\left\{\frac{\partial f_n(\mathbf{y})}{\partial y_n}\right\}$$

Since the expectation is a linear operator, (2.3) follows directly. ■

By applying Lemma 1 to the expression of the MSE, we then get Stein's Unbiased Risk (or expected MSE) Estimate (SURE):

Theorem 1. *Under the same hypotheses as Lemma 1, the random variable*

$$\epsilon = \frac{1}{N}\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + \frac{2\sigma^2}{N}\mathrm{div}\{\mathbf{F}(\mathbf{y})\} - \sigma^2 \tag{2.4}$$

is an unbiased estimator of the expected MSE, i.e.

$$\mathscr{E}\{\epsilon\} = \frac{1}{N}\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y}) - \mathbf{x}\|^2\right\}$$

Proof. By expanding the expectation of the MSE, we have that

$$\begin{aligned}
\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y}) - \mathbf{x}\|^2\right\} &= \mathscr{E}\left\{\|\mathbf{F}(\mathbf{y})\|^2\right\} - 2\mathscr{E}\left\{\mathbf{F}(\mathbf{y})^{\mathrm{T}}\mathbf{x}\right\} + \mathscr{E}\left\{\|\mathbf{x}\|^2\right\} \\
&= \mathscr{E}\left\{\|\mathbf{F}(\mathbf{y})\|^2\right\} - 2\mathscr{E}\left\{\mathbf{F}(\mathbf{y})^{\mathrm{T}}\mathbf{y}\right\} + 2\sigma^2\mathscr{E}\{\mathrm{div}\{\mathbf{F}(\mathbf{y})\}\} \\
&\quad + \|\mathbf{x}\|^2
\end{aligned}$$

[3]The integrated term $[f_n(\mathbf{y})q(y_n - x_n)]_{-\infty}^{+\infty}$ vanishes by hypothesis.

where we have applied Lemma 1. Using the statistical independence between $\mathbf{x}$ and $\mathbf{b}$, we can estimate the constant energy term $\|\mathbf{x}\|^2$ as

$$\begin{aligned}\|\mathbf{x}\|^2 &= \mathscr{E}\left\{\|\mathbf{y}-\mathbf{b}\|^2\right\} \\ &= \mathscr{E}\left\{\|\mathbf{y}\|^2\right\} + \underbrace{\mathscr{E}\left\{\|\mathbf{b}\|^2\right\}}_{N\sigma^2} - 2\underbrace{\mathscr{E}\left\{\mathbf{y}^{\mathrm{T}}\mathbf{b}\right\}}_{\mathscr{E}\{\mathbf{x}^{\mathrm{T}}\mathbf{b}\}+\mathscr{E}\{\|\mathbf{b}\|^2\}} \\ &= \mathscr{E}\left\{\|\mathbf{y}\|^2\right\} - N\sigma^2\end{aligned}$$

A rearrangement of the $\mathbf{y}$-terms then provides the result of Theorem 1. ■

Note that the estimate of the constant energy term $\|\mathbf{x}\|^2$ is irrelevant to the SURE-based optimization of the denoising process $\mathbf{F}$.

In the following propositions, we derive two natural extensions to the standard SURE result given in Theorem 1, as well as a more general way of of obtaining Stein-like lemmas.

Proposition 1. *Let $\mathbf{W}$ be an arbitrary $N \times N$ matrix and define the weighted norm $\|\cdot\|_{\mathbf{W}}$ as $\|\mathbf{y}\|_{\mathbf{W}} = \|\mathbf{W}\mathbf{y}\|_2$. Then, under the same hypothesis as Lemma 1, the following random variable*

$$\epsilon = \frac{1}{N}\left(\|\mathbf{F}(\mathbf{y})-\mathbf{y}\|_{\mathbf{W}}^2 + 2\sigma^2 \mathrm{div}\left\{\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{F}(\mathbf{y})\right\} - \sigma^2 \mathrm{trace}\left\{\mathbf{W}\mathbf{W}^{\mathrm{T}}\right\}\right) \tag{2.5}$$

is an unbiased estimate of the expected weighted MSE, i.e.

$$\mathscr{E}\{\epsilon\} = \frac{1}{N}\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y})-\mathbf{x}\|_{\mathbf{W}}^2\right\}$$

Proof. We can expand the expected weighted squared error as

$$\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y})-\mathbf{x}\|_{\mathbf{W}}^2\right\} = \mathscr{E}\left\{\|\mathbf{F}(\mathbf{y})\|_{\mathbf{W}}^2\right\} + \underbrace{\|\mathbf{x}\|_{\mathbf{W}}^2}_{\text{(I)}} - 2\underbrace{\mathscr{E}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{F}(\mathbf{y})\right\}}_{\text{(II)}}$$

We can then further develop the two quantities (I) and (II) which involve the unknown data $\mathbf{x}$:

$$\begin{aligned}\text{(I)}\quad \|\mathbf{x}\|_{\mathbf{W}}^2 &= \mathscr{E}\left\{\|\mathbf{y}-\mathbf{b}\|_{\mathbf{W}}^2\right\} \\ &= \mathscr{E}\left\{\|\mathbf{y}\|_{\mathbf{W}}^2\right\} - \mathscr{E}\left\{\|\mathbf{b}\|_{\mathbf{W}}^2\right\} \\ &= \mathscr{E}\left\{\|\mathbf{y}\|_{\mathbf{W}}^2\right\} - \mathrm{trace}\left\{\mathbf{W}\mathscr{E}\left\{\mathbf{b}\mathbf{b}^{\mathrm{T}}\right\}\mathbf{W}^{\mathrm{T}}\right\} \\ &= \mathscr{E}\left\{\|\mathbf{y}\|_{\mathbf{W}}^2\right\} - \sigma^2\mathrm{trace}\left\{\mathbf{W}\mathbf{W}^{\mathrm{T}}\right\}\end{aligned}$$

$$\begin{aligned}\text{(II)}\quad \mathscr{E}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{F}(\mathbf{y})\right\} &= \sum_{i,j,k} w_{i,j}w_{i,k}\mathscr{E}\left\{x_j f_k(\mathbf{y})\right\} \\ &\overset{\text{Lemma 1}}{=} \sum_{i,j,k} w_{i,j}w_{i,k}\left(\mathscr{E}\left\{y_j f_k(\mathbf{y})\right\} - \sigma^2\mathscr{E}\left\{\frac{\partial f_k(\mathbf{y})}{\partial y_j}\right\}\right) \\ &= \mathscr{E}\left\{\mathbf{y}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{F}(\mathbf{y})\right\} - \sigma^2\mathrm{div}\left\{\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{F}(\mathbf{y})\right\}\end{aligned}$$

A rearrangement of the $\mathbf{y}$-terms finally provides the desired result. ■

Proposition 2 (Generalization of Stein's Lemma)**.** *Let $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$ be an N-dimensional Gaussian random vector and $\boldsymbol{\lambda} \in \mathbb{R}^N$ an arbitrary N-dimensional deterministic vector. Then, for any $k \in \mathbb{N}$, we have the following equality:*

$$\mathscr{E}\left\{b_n^k f_n(\mathbf{y})\right\} = \mathscr{E}\left\{\left[\frac{\partial^k}{\partial \lambda_n^k}\left(\exp\Big(\frac{\|\boldsymbol{\lambda}\|^2_{\mathbf{\Gamma}^{1/2}}}{2}\Big) f_n(\mathbf{y}+\mathbf{\Gamma}\boldsymbol{\lambda})\right)\right]_{\boldsymbol{\lambda}=\mathbf{0}}\right\} \tag{2.6}$$

Proof. We begin by noticing that:

$$\begin{aligned}\mathscr{E}\left\{b_n^k f_n(\mathbf{y})\right\} &= \mathscr{E}\left\{\left[\frac{\partial^k}{\partial \lambda_n^k}\exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}) f_n(\mathbf{y})\right]_{\boldsymbol{\lambda}=\mathbf{0}}\right\} \\ &= \left[\frac{\partial^k}{\partial \lambda_n^k}\mathscr{E}\left\{\exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}) f_n(\mathbf{y})\right\}\right]_{\boldsymbol{\lambda}=\mathbf{0}} \end{aligned} \tag{2.7}$$

To be fully rigorous, one has to assume that $\mathscr{E}\left\{\left|\frac{\partial^k}{\partial \lambda_n^k}\exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}) f_n(\mathbf{y})\right|_{\boldsymbol{\lambda}=\mathbf{0}}\right\} < +\infty$ to switch the expectation and the partial derivative operators.

Since $\mathbf{b}$ follows a zero-mean multivariate Gaussian distribution with covariance matrix $\mathbf{\Gamma}$, its probability density function is given by:

$$q(\mathbf{b}) = \mathrm{cst}\cdot\exp\Big(-\frac{\mathbf{b}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\mathbf{b}}{2}\Big) = \mathrm{cst}\cdot\exp\Big(-\frac{\|\mathbf{b}\|^2_{\mathbf{\Gamma}^{-1/2}}}{2}\Big)$$

Consequently, we have the following equality:

$$\begin{aligned}\exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b})q(\mathbf{b}) &= \mathrm{cst}\cdot\exp\Big(-\frac{1}{2}(\mathbf{b}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\mathbf{b} - 2\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b} + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{\Gamma}\boldsymbol{\lambda}) + \frac{\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{\Gamma}\boldsymbol{\lambda}}{2}\Big) \\ &= \exp\Big(\frac{\|\boldsymbol{\lambda}\|^2_{\mathbf{\Gamma}^{1/2}}}{2}\Big)\cdot\underbrace{\mathrm{cst}\cdot\exp\Big(-\frac{\|\mathbf{b}-\mathbf{\Gamma}\boldsymbol{\lambda}\|^2_{\mathbf{\Gamma}^{-1/2}}}{2}\Big)}_{q(\mathbf{b}-\mathbf{\Gamma}\boldsymbol{\lambda})} \end{aligned} \tag{2.8}$$

We can then successively write:

$$\begin{aligned}\mathscr{E}\left\{\exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}) f_n(\mathbf{y})\right\} &= \int_{\mathbb{R}^N} f_n(\mathbf{y})\exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b})q(\mathbf{b})\mathrm{d}\mathbf{b} \\ &\overset{(2.8)}{=} \exp\Big(\frac{\|\boldsymbol{\lambda}\|^2_{\mathbf{\Gamma}^{1/2}}}{2}\Big)\int_{\mathbb{R}^N} f_n(\mathbf{y})q(\mathbf{b}-\mathbf{\Gamma}\boldsymbol{\lambda})\mathrm{d}\mathbf{b} \\ &= \exp\Big(\frac{\|\boldsymbol{\lambda}\|^2_{\mathbf{\Gamma}^{1/2}}}{2}\Big)\int_{\mathbb{R}^N} f_n(\mathbf{y}+\mathbf{\Gamma}\boldsymbol{\lambda})q(\mathbf{b})\mathrm{d}\mathbf{b} \\ &= \mathscr{E}\left\{\exp\Big(\frac{\|\boldsymbol{\lambda}\|^2_{\mathbf{\Gamma}^{1/2}}}{2}\Big) f_n(\mathbf{y}+\mathbf{\Gamma}\boldsymbol{\lambda})\right\} \end{aligned} \tag{2.9}$$

The desired result (2.6) is finally demonstrated by substituting (2.9) into (2.7). ■

Stein's lemma, i.e. $\mathscr{E}\{b_n f_n(\mathbf{y})\} = \sigma^2 \mathscr{E}\{\partial f_n(\mathbf{y})/\partial y_n\}$ can be directly obtained from Proposition 2, by considering $\mathbf{\Gamma} = \sigma^2 \mathbf{Id}$ and $k = 1$.

Proposition 3. *In the case of additive* colored *Gaussian noise, i.e.* $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$, *the unbiased estimator of the MSE given in Theorem 1 becomes:*

$$\epsilon = \frac{1}{N} \left(\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + 2 \operatorname{div} \{\mathbf{\Gamma F}(\mathbf{y})\} - \operatorname{trace} \{\mathbf{\Gamma}\} \right) \tag{2.10}$$

where $\mathbf{\Gamma}$ *is the* $N \times N$ *covariance matrix of the colored noise* $\mathbf{b}$.

Proof. Using the fact that $\mathbf{x} = \mathbf{y} - \mathbf{b}$, as well as the standard result $\mathscr{E}\left\{\|\mathbf{y}\|^2\right\} = \|\mathbf{x}\|^2 + \operatorname{trace}\{\mathbf{\Gamma}\}$ allows us to rewrite the expected MSE as:

$$\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y}) - \mathbf{x}\|^2\right\} = \mathscr{E}\left\{\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2\right\} + 2\mathscr{E}\left\{\mathbf{b}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\right\} - \operatorname{trace}\{\mathbf{\Gamma}\}$$

The result of Proposition 2 (for $k = 1$) can then be used to evaluate the second expectation as:

$$\begin{aligned} \mathscr{E}\left\{\mathbf{b}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\right\} &= \left[\mathscr{E}\left\{\operatorname{div}_{\boldsymbol{\lambda}}\{\exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b})\mathbf{F}(\mathbf{y})\}\right\}\right]_{\boldsymbol{\lambda}=\mathbf{0}} \\ &= \left[\mathscr{E}\left\{\operatorname{div}_{\boldsymbol{\lambda}}\{\exp\left(\frac{\|\boldsymbol{\lambda}\|^2_{\mathbf{\Gamma}^{1/2}}}{2}\right)\mathbf{F}(\mathbf{y} + \mathbf{\Gamma}\boldsymbol{\lambda})\}\right\}\right]_{\boldsymbol{\lambda}=\mathbf{0}} \\ &= \mathscr{E}\{\operatorname{div}\{\mathbf{\Gamma F}(\mathbf{y})\}\} \end{aligned}$$

which finally leads to the desired result (2.10). ■

2.4 Reliability of the SURE Estimate

In the previous section, we showed that SURE is a random variable that has the same expectation as the MSE. We now evaluate its reliability by computing the expected squared error between SURE and the actual MSE[5]. For the sake of simplicity, we consider the AWGN model described in (2.1) and we do not take into account the error induced by the estimation of the noise-free signal energy $\|\mathbf{x}\|^2$, because this term does not appear in the minimization of SURE/MSE.

Property 1. *In the case of AWGN, the expected squared error between the effective part of SURE* (2.4) *and the actual MSE* (2.2) *is given by:*

$$\mathscr{E}\left\{(\epsilon - \mathrm{MSE})^2\right\} = \frac{4\sigma^2}{N^2}\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y})\|^2 + \sigma^2 \operatorname{trace}\left\{\mathbf{J}_{\mathbf{F}}(\mathbf{y})^2\right\}\right\} \tag{2.11}$$

where $\mathbf{J}_{\mathbf{F}}(\mathbf{y}) = \left[\frac{\partial f_i(\mathbf{y})}{\partial y_j}\right]_{1 \le i,j \le N}$ *is the Jacobian matrix of the vector-valued function* $\mathbf{F}(\mathbf{y})$.

[5]Note that a study on the reliability of SURE as an estimate of the MSE for applications in deconvolution has been recently proposed by Pesquet in [85].

Proof. By definition of the MSE, we can expand the expected squared error as:

$$\begin{aligned}\mathscr{E}\left\{(\epsilon - \text{MSE})^2\right\} &= \frac{4}{N^2}\mathscr{E}\Big\{\Big(\sigma^2 \text{div}\{\mathbf{F}(\mathbf{y}) - (\underbrace{\mathbf{y}-\mathbf{x}}_{\mathbf{b}})^{\mathrm{T}}\mathbf{F}(\mathbf{y})\}\Big)^2\Big\} \\ &= \frac{4}{N^2}\Big(\sigma^4\mathscr{E}\left\{\text{div}\left\{\mathbf{F}(\mathbf{y})\right\}^2\right\} - 2\sigma^2 \underbrace{\mathscr{E}\left\{\text{div}\left\{\mathbf{F}(\mathbf{y})\right\}\mathbf{b}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\right\}}_{\text{(I)}} \\ &\quad + \underbrace{\mathscr{E}\left\{(\mathbf{b}^{\mathrm{T}}\mathbf{F}(\mathbf{y}))^2\right\}}_{\text{(II)}}\Big) \end{aligned} \tag{2.12}$$

Only the expectations (I) and (II) contain the unknown noise component $\mathbf{b}$. Using integration by parts, we rewrite them as:

$$\begin{aligned}\text{(I)}\quad \mathscr{E}\left\{\text{div}\left\{\mathbf{F}(\mathbf{y})\right\}\mathbf{b}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\right\} &= \sum_{i,j}\mathscr{E}\left\{\frac{\partial f_i(\mathbf{y})}{\partial y_i}b_j f_j(\mathbf{y})\right\} \\ &= \sigma^2\sum_{i,j}\mathscr{E}\left\{\frac{\partial f_i(\mathbf{y})}{\partial y_i}\frac{\partial f_j(\mathbf{y})}{\partial y_j} + \frac{\partial^2 f_i(\mathbf{y})}{\partial y_i \partial y_j}f_j(\mathbf{y})\right\} \\ \text{(II)}\quad \mathscr{E}\left\{(\mathbf{b}^{\mathrm{T}}\mathbf{F}(\mathbf{y}))^2\right\} &= \sum_{i,j}\mathscr{E}\left\{b_i b_j f_i(\mathbf{y}) f_j(\mathbf{y})\right\} \\ &= \sigma^2\sum_{i}\mathscr{E}\left\{f_i^2(\mathbf{y})\right\} + \sigma^4\sum_{i,j}\mathscr{E}\left\{\frac{\partial f_i(\mathbf{y})}{\partial y_j}\frac{\partial f_j(\mathbf{y})}{\partial y_i}\right\} \\ &\quad + \sigma^4\sum_{i,j}\mathscr{E}\left\{2\frac{\partial^2 f_i(\mathbf{y})}{\partial y_i \partial y_j}f_j(\mathbf{y}) + \frac{\partial f_i(\mathbf{y})}{\partial y_i}\frac{\partial f_j(\mathbf{y})}{\partial y_j}\right\}\end{aligned}$$

where we have used the result of Proposition 2 to derive

$$\mathscr{E}\left\{b_i^2 f_i^2(\mathbf{y})\right\} = \sigma^2\mathscr{E}\left\{f_i^2(\mathbf{y})\right\} + \sigma^4\mathscr{E}\left\{\frac{\partial^2 f_i^2(\mathbf{y})}{\partial y_i^2}\right\}.$$

Now computing:

$$\begin{aligned}\text{(II)} - 2\sigma^2\text{(I)} &= \sigma^2\sum_{i}\mathscr{E}\left\{f_i^2(\mathbf{y})\right\} + \sigma^4\sum_{i,j}\mathscr{E}\left\{\frac{\partial f_i(\mathbf{y})}{\partial y_j}\frac{\partial f_j(\mathbf{y})}{\partial y_i}\right\} \\ &\quad - \sigma^4\sum_{i,j}\mathscr{E}\left\{\frac{\partial f_i(\mathbf{y})}{\partial y_i}\frac{\partial f_j(\mathbf{y})}{\partial y_j}\right\} \\ &= \sigma^2\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y})\|^2\right\} + \sigma^4\mathscr{E}\left\{\text{trace}\left\{\mathbf{J}_{\mathbf{F}}(\mathbf{y})^2\right\}\right\} \\ &\quad - \sigma^4\mathscr{E}\left\{\text{div}\left\{\mathbf{F}(\mathbf{y})\right\}^2\right\}\end{aligned} \tag{2.13}$$

and putting back (2.13) into (2.12) finally leads to the desired result (2.11). ■

As it can be observed in Equ. (2.11), the squared error between SURE and MSE depends on three parameters: the number of samples N, the variance of the AWGN and the design of the denoising process. As an illustrative example, we propose to derive an upper bound of this squared error in the particular case where $\mathbf{F}$ is a *linear* processing, i.e. $\mathbf{F}(\mathbf{y}) = \mathbf{A}\mathbf{y}$, where $\mathbf{A}$ is a $N \times N$ matrix. We further assume that $\mathbf{A}$ can be diagonalized in an orthonormal basis, i.e. $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{H}}$, where $\mathbf{U}\mathbf{U}^{\mathrm{H}} = \mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{Id}$. In denoising, the diagonal matrix $\mathbf{S} = [s_{i,j}]_{N \leq i,j \leq N}$ can be interpreted as a *pointwise* shrinkage, i.e. $0 \leq s_{i,i} \leq 1$, $i = 1 \dots N$. We denote the transformed coefficients by $\mathbf{w} = \mathbf{U}^{\mathrm{H}}\mathbf{y}$ and we have $\|\mathbf{w}\|^2 = \|\mathbf{y}\|^2$. In the considered framework, we can derive the following upper bounds:

1. $\|\mathbf{F}(\mathbf{y})\|^2 = \mathbf{y}^{\mathrm{T}}\mathbf{U}\mathbf{S}^2\mathbf{U}^{\mathrm{H}}\mathbf{y} = \mathbf{w}^{\mathrm{H}}\mathbf{S}^2\mathbf{w} \leq \|\mathbf{y}\|^2$
2. $\mathrm{trace}\left\{\mathbf{J}_{\mathbf{F}}(\mathbf{y})^2\right\} = \mathrm{trace}\left\{\mathbf{A}^2\right\} = \mathrm{trace}\left\{\mathbf{U}\mathbf{S}^2\mathbf{U}^{\mathrm{H}}\right\} = \mathrm{trace}\left\{\mathbf{S}^2\right\} \leq N$

and thus:

$$\mathscr{E}\left\{(\epsilon - \mathrm{MSE})^2\right\} \leq \frac{4\sigma^2}{N^2}\left(\mathscr{E}\left\{\|\mathbf{y}\|^2\right\} + N\sigma^2\right)$$

Now introducing the mean $\mu_{\mathbf{y}}$ and the variance $\sigma^2_{\mathbf{y}}$ of the noisy data $\mathbf{y}$ defined as:

$$\left\{\begin{array}{rcl} \mu_{\mathbf{y}} & = & \frac{1}{N}\mathscr{E}\left\{\mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} \\ \sigma^2_{\mathbf{y}} & = & \frac{1}{N}\mathscr{E}\left\{\|\mathbf{y}\|^2\right\} - \mu^2_{\mathbf{y}} \end{array}\right.$$

allows us to finally conclude that:

$$\mathscr{E}\left\{(\epsilon - \mathrm{MSE})^2\right\} \leq \frac{4\sigma^2}{N}\left(\sigma^2_{\mathbf{y}} + \mu^2_{\mathbf{y}} + \sigma^2\right)$$

In the particular case of a *pointwise* shrinkage in an *orthonormal* transform domain, the squared error between SURE and MSE thus decreases as the number of samples N increases.

2.5 Optimal Estimator

When looking at the MSE estimate given in (2.4), a natural question arises: which processing $\mathbf{F}$ of the observed data $\mathbf{y}$, is the minimizer of the expected MSE ? To answer this question, one has to assume that the underlying noise-free signal $\mathbf{x}$ is a random variable with probability density function (PDF) $p(\mathbf{x})$. In that case, the PDF of the observed data $\mathbf{y}$ is the convolution between $p(\mathbf{x})$ and the PDF of the AWGN $q(\mathbf{b})$, i.e.

$$\mathbf{y} \sim r(\mathbf{y}) = (p * q)(\mathbf{y}) \tag{2.14}$$

We can then state the following theorem[6]:

[6] A similar result can be found in [70, 86].

Theorem 2. *The optimal, in the minimum expected MSE sense, estimate of a signal degraded by an AWGN of variance σ^2 is given by:*

$$\mathbf{F}_{\text{opt}}(\mathbf{y}) = \mathbf{y} + \sigma^2 \nabla \log r(\mathbf{y}) \tag{2.15}$$

where $r(\mathbf{y})$ is the probability density function of the observed noisy data $\mathbf{y}$.

Proof. Taking the expectation of SURE (which is equivalent to the expectation of the MSE) w.r.t. $\mathbf{y}$ leads to:

$$\begin{aligned} N\mathscr{E}_{\mathbf{y}}\{\epsilon\} &= \int_{\mathbb{R}^N} \left(\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + 2\sigma^2 \text{div}\left\{\mathbf{F}(\mathbf{y})\right\} \right) r(\mathbf{y}) \text{d}\mathbf{y} - N\sigma^2 \\ &= \sum_{n=1}^{N} \left(\int_{\mathbb{R}^N} (f_n(\mathbf{y}) - y_n)^2 r(\mathbf{y}) \text{d}\mathbf{y} + 2\sigma^2 \int_{\mathbb{R}^N} \frac{\partial f_n(\mathbf{y})}{\partial y_n} r(\mathbf{y}) \text{d}\mathbf{y} \right) - N\sigma^2 \end{aligned}$$

Assuming $\lim_{\mathbf{y} \to \pm\infty} f_n(\mathbf{y}) r(\mathbf{y}) = 0$, the integration by parts of the second integral gives:

$$\int_{\mathbb{R}^N} \frac{\partial f_n(\mathbf{y})}{\partial y_n} r(\mathbf{y}) \text{d}\mathbf{y} = - \int_{\mathbb{R}^N} f_n(\mathbf{y}) \frac{\partial r(\mathbf{y})}{\partial y_n} \text{d}\mathbf{y}$$

The expectation of SURE can be thus rewritten as:

$$N\mathscr{E}_{\mathbf{y}}\{\epsilon\} = \sum_{n=1}^{N} \underbrace{\int_{\mathbb{R}^N} \left((f_n(\mathbf{y}) - y_n)^2 r(\mathbf{y}) - 2\sigma^2 f_n(\mathbf{y}) \frac{\partial r(\mathbf{y})}{\partial y_n} \right) \text{d}\mathbf{y}}_{J(f_n)} - N\sigma^2 \tag{2.16}$$

Now denoting by f_n^{opt} the global optimum of the quadratic functional $J(f_n)$ and considering an arbitrary function $h(\mathbf{y}) \in C_0^\infty(\mathbb{R}^N)$[7], we must have $\forall n \in [1; N]$ and $t \in \mathbb{R}$,

$$\begin{aligned} \frac{\text{d}}{\text{d}t} J(f_n^{\text{opt}} + th)\Big|_{t=0} &= 0 \\ &\Updownarrow \\ \int_{\mathbb{R}^N} h(\mathbf{y}) \left((f_n^{\text{opt}}(\mathbf{y}) - y_n) r(\mathbf{y}) - \sigma^2 \frac{\partial r(\mathbf{y})}{\partial y_n} \right) \text{d}\mathbf{y} &= 0 \end{aligned} \tag{2.17}$$

A fundamental lemma of the calculus of variations[8] allows us to conclude that the equality (2.17) implies

$$(f_n^{\text{opt}}(\mathbf{y}) - y_n) r(\mathbf{y}) - \sigma^2 \frac{\partial r(\mathbf{y})}{\partial y_n} = 0$$

[7] Here, $C_0^\infty(\mathbb{R}^N)$ is the space of all infinitely differentiable functions defined on $\mathbb{R}^N$ whose support is a compact set contained in $\mathbb{R}^N$.

[8] Known as the "du Bois-Reymond" lemma [87].

and thus $\forall n \in [1;N]$, $f_n^{\text{opt}}(\mathbf{y}) = y_n + \sigma^2 \dfrac{\partial}{\partial y_n} \log r(\mathbf{y})$, which finally leads to the desired result:

$$\mathbf{F}_{\text{opt}}(\mathbf{y}) = \mathbf{y} + \sigma^2 \nabla \log r(\mathbf{y})$$

■

The result of Theorem 2 can be extended to additive colored noise $\mathbf{b} \sim \mathcal{N}(0, \mathbf{\Gamma})$. In that case, the optimal denoising function is given by:

$$\mathbf{F}_{\text{opt}}(\mathbf{y}) = \mathbf{y} + \mathbf{\Gamma} \nabla \log r(\mathbf{y}) \tag{2.18}$$

The use of the optimal estimator requires the evaluation of the PDF of the noisy observations. Two strategies are then conceivable:

1. **Prior-free approach:** In this case, the PDF is directly estimated from the data. This can be achieved by interpolating the histogram of the noisy samples or approximating the PDF by linear combination of smoothing kernels centered around each sample (Parzen method [88]). There are numerous other methods for estimating PDF; some of them are reviewed in [89]. The estimated PDF has to be very accurate, because small errors on the PDF estimation might yield large errors on the derivative of its logarithm, and consequently lead to unreliable denoising function. While estimating PDF is reasonably feasible for univariate PDF, it becomes much more involved for multivariate PDF.

2. **Bayesian approach:** In the Bayesian framework, the optimal (in the MMSE sense) estimator is known as the Bayesian least-square, i.e.

   $$\mathbf{F}_{\text{opt}}(\mathbf{y}) = \mathscr{E}\{\mathbf{x}|\mathbf{y}\}$$

   To derive its explicit expression, a prior on the PDF of the underlying noise-free signal is formulated. As recalled at the beginning of this section, its convolution with the noise PDF gives the PDF of the noisy data (2.14). For instance, let us assume that, in a given transform-domain, $\mathbf{x} \sim \mathcal{N}(0, \mathbf{\Gamma_x})$ and $\mathbf{b} \sim \mathcal{N}(0, \mathbf{\Gamma_b})$. Then the PDF of the noisy data is given by $r(\mathbf{y}) = \text{cst} \times \exp\left(-\frac{1}{2}\mathbf{y}^{\text{T}}(\mathbf{\Gamma_x} + \mathbf{\Gamma_b})^{-1}\mathbf{y}\right)$ and the corresponding optimal processing is nothing but the classical Wiener filter that we briefly discussed in Section 1.3.1:

   $$\mathbf{F}_{\text{opt}}(\mathbf{y}) = \mathbf{\Gamma_x}(\mathbf{\Gamma_x} + \mathbf{\Gamma_b})^{-1}\mathbf{y}$$

   There are two main drawbacks with the Bayesian approach. First, it might be difficult to derive an analytical expression for $\mathbf{F}(\mathbf{y})$ when assuming a sophisticated prior on the noise-free data. Moreover, such prior usually involves some non-linear parameters that have to be numerically optimized based on the observed noisy data. Second, if the prior does not properly reflect the

statistics of the noise-free data, the resulting denoising function will not be efficient in removing the noise component and/or preserving the underlying noise-free signal.

Nevertheless, as shown in Theorem 2, the derivation of the so-called Bayesian least-square estimator does not require the explicit statistical modeling of the underlying noise-free signal. This observation has been recently emphasized by Raphan and Simoncelli [86], who consider SURE as a "prior-free" Bayesian approach.

2.6 Linear Expansion of Thresholds (LET)

In the previous section, we have pointed out that the optimal minimum MSE estimator is difficult to implement in practice, because it requires an accurate estimation of the PDF of the noisy data. In this section, we adopt another point of view: instead of parametrizing the noisy data PDF and then deriving the corresponding estimator, we directly parametrize the estimator. More precisely, we propose to describe the denoising process $\mathbf{F}$ as a *linear* combination of K possibly non-linear processing $\mathbf{F}_k$, i.e.

$$\mathbf{F}(\mathbf{y}) = \sum_{k=1}^{K} a_k \mathbf{F}_k(\mathbf{y}) \tag{2.19}$$

Since the effective part of the processing is usually a thresholding performed in a transform-domain, we have coined the above formulation of the denoising process a *linear expansion of thresholds* (LET). As in every estimation problems, the number of parameters K used to describe the estimator $\mathbf{F}(\mathbf{y})$ must remain much lower than the number of observations N to preserve the "smoothing" effect of $\mathbf{F}$ and avoid fitting of the noisy data.

The LET approach has two main advantages. First, it does not require any statistical assumptions on the unknown noise-free data in order to design the thresholding functions. Second, the search for the optimal (in the minimum SURE/MSE sense) parameters a_k's boils down to the solution of a *linear* system of equations, thanks to the quadratic form of the unbiased MSE estimate. Indeed, introducing Equ. (2.19) into the result of Theorem 1 and performing differentiation over a_k leads to

$$\begin{aligned}
\frac{N}{2}\frac{\partial \epsilon}{\partial a_k} &= \sum_{l=1}^{K} \underbrace{\mathbf{F}_k(\mathbf{y})^{\mathrm{T}}\mathbf{F}_l(\mathbf{y})}_{[\mathbf{M}]_{k,l}} a_l - \Big(\underbrace{\mathbf{F}_k(\mathbf{y})^{\mathrm{T}}\mathbf{y} - \sigma^2 \mathrm{div}\left\{\mathbf{F}_k(\mathbf{y})\right\}}_{[\mathbf{c}]_k} \Big) \\
&= 0, \; \forall k \in [1;K] \\
&\Updownarrow \\
\mathbf{Ma} &= \mathbf{c}
\end{aligned} \tag{2.20}$$

Since the minimum of ϵ always exists, we are ensured that there will always be a solution to the above linear system. When $\text{rank}(\mathbf{M}) < K$, the function $\mathbf{F}$ is *over-parameterized* and consequently, several sets of parameters a_k yield equivalent minimum SURE results. In that case, we can simply consider the solution provided by the pseudoinverse of $\mathbf{M}$. Of course, it is also possible to reduce the parametrization order K so as to make the matrix $\mathbf{M}$ full rank (at no SURE quality loss).

Note that, in the context of image denoising, a *univariate* linear parameterization combined with an *implicit* SURE minimization was already evoked in [72] (*sigmoidal filtering*), although not much promoted.

What this SURE-LET approach suggests is that the practitioner may choose at will (restricted only by the differentiability constraint of Theorem 1) a set of K different denoising algorithms (ideally with complementary denoising behaviors) and optimize a weighting of these algorithms to get the best of them at once.

Among the potentially interesting algorithms are those that work in a transformed domain such as

- the non-redundant wavelet transforms, either orthogonal or bi-orthogonal [38, 39];
- the classical undecimated wavelet transform [90];
- the *curvelet* [52] transform;
- the *contourlet* [55] transform;
- the steerable pyramids [46, 48];
- the discrete cosine transform (DCT) or its overcomplete variant: the overlapping block discrete cosine transform (BDCT).

In the two forthcoming chapters, we will further discuss the application of the SURE-LET paradigm to some transformed domain denoising.

2.7 Summary

In this chapter, we have devised a general procedure (SURE-LET) for denoising images corrupted by AWGN. The two key ingredients of the proposed approach are:

1. An *unbiased* estimate of the MSE, known as SURE. In particular, this statistical quantity only depends on the observed noisy measurements and can thus be computed in practice. In contrast to the popular Bayesian approach, no prior on the unknown noise-free data is required to derive SURE.

2. A linear expansion of thresholds (LET): in image denoising, SURE was usually used for optimizing one or several *non-linear* parameters involved in ad hoc wavelet estimators. In this chapter, we have proposed an alternative approach by building a *linearly* parametrized denoiser which offers more flexibility than the standard thresholding functions. Thanks to the quadratic form of SURE, we have shown that the optimal (in the minimum SURE sense) parameters are simply the solution of a *linear* system of equations. From a computational point of view, this makes the SURE-LET strategy particularly attractive.

Chapter 3

SURE-LET in Orthonormal Wavelet Representation

3.1 Motivations

Despite reports on the superior denoising performances of redundant transforms [43, 52, 91], we will only consider critically sampled wavelet transforms in this chapter[1]. The rationale behind our choice is that, since there is no *added* information (only *repeated* information) in redundant transforms, we believe that, eventually, a non-redundant transform may match the performance of redundant ones. This would potentially be very promising since the major drawback of redundant transforms are their memory and CPU time requirements which limits their routine use for very large images and, above all, usual volumes of data encountered in biomedical applications.

Besides its lighter computational cost, an orthonormal wavelet transform (OWT) has two further advantages over redundant transformations:

- **Energy conservation:** The Euclidian norm is preserved in the transformed domain. In particular, the mean-squared error (MSE) in the image domain is a weighted sum of the MSEs of each individual orthonormal wavelet subband $\mathbf{x}^j \in \mathbb{R}^{N_j}$:

$$\underbrace{\frac{1}{N}\|\hat{\mathbf{x}} - \mathbf{x}\|^2}_{\text{MSE}} = \sum_{j=1}^{J+1} \frac{N_j}{N} \underbrace{\|\hat{\mathbf{x}}^j - \mathbf{x}^j\|^2}_{\text{MSE}^j} \tag{3.1}$$

  where N_j is the number of pixels in subband $j = 1 \dots J+1$.

[1] Some parts of this chapter are based on our published papers [92, 93].

- **Preservation of the AWGN model:** The noise remains white and Gaussian with same statistics in the orthonormal wavelet domain. From the noise point of view, the wavelet subbands are therefore statistically independent, and consequently

$$\mathscr{E}\left\{\mathbf{b}^i \mathbf{b}^{j\,\mathrm{T}}\right\} = \sigma^2 \mathbf{Id}\, \delta_{i-j} \tag{3.2}$$

where $\delta_n = \begin{cases} 1, & \text{if } n = 0 \\ 0, & \text{otherwise} \end{cases}$ is the discrete Kronecker delta function.

In each orthonormal wavelet subband $j \in [1; J+1]$, we thus have the following observation model:

$$\mathbf{y}^j = \mathbf{x}^j + \mathbf{b}^j, \text{ with } \mathbf{b}^j \sim \mathcal{N}(0, \sigma^2 \mathbf{Id}) \tag{3.3}$$

In the orthonormal wavelet domain, each subband can therefore be denoised independently.

3.2 Pointwise Estimator

In this section, we only consider *pointwise*, subband-adaptive wavelet estimators of the form

$$\hat{\mathbf{x}}^j = [\theta^j(y_n^j)]_{1 \le N_j}, \text{ for } j \in [1; J] \tag{3.4}$$

As it is usually the case in wavelet denoising, the lowpass residual subband $\mathbf{y}^{J+1}$ is not processed, i.e. $\hat{\mathbf{x}}^{J+1} = \mathbf{y}^{J+1}$. From now on, we will drop the subband index j since a new denoising function is independently applied in each individual subband.

3.2.1 Pointwise SURE

Inside each subband of an orthonormal wavelet transform, the MSE is given by

$$\mathrm{MSE} = \frac{1}{N} \sum_{n=1}^{N} \left(\theta(y_n) - x_n\right)^2 \tag{3.5}$$

In Section 2.3, we gave a general form of Stein's unbiased risk estimate for an arbitrary processing. Below, we re-express SURE for an orthonormal wavelet-domain *pointwise* processing θ.

Corollary 1. *Let $\theta : \mathbb{R} \to \mathbb{R}$ be a (weakly) differentiable function, such that $|\theta(z)| \le cste \cdot \exp(az^2)$ for $a < \frac{1}{2\sigma^2}$. Then, the following random variable:*

$$\epsilon \;=\; \frac{1}{N} \sum_{n=1}^{N} \left((\theta(y_n) - y_n)^2 + 2\sigma^2 \theta'(y_n)\right) - \sigma^2 \tag{3.6}$$

is an unbiased estimator of the expected MSE, i.e.

$$\mathscr{E}\{\epsilon\} = \mathscr{E}\left\{\frac{1}{N}\sum_{n=1}^{N}(\theta(y_n) - x_n)^2\right\}$$

Proof. Since this corollary is a particular case of Theorem 1, it is admitted without proof. ■

3.2.2 Pointwise SURE-LET

In Section 2.6, we have proposed a general form of denoising functions (2.19): a *linear expansion of thresholds* (LET). In the context of pointwise orthonormal wavelet denoising, LET takes the following form:

$$\theta(y_n) = \sum_{k=1}^{K} a_k \theta_k(y_n) \tag{3.7}$$

The vector of parameters $\mathbf{a} = [a_1 \dots a_K]^{\mathrm{T}}$ that minimizes (3.6) for the above pointwise LET is given by:

$$\mathbf{a}_{\mathrm{opt}} = \mathbf{M}^{-1}\mathbf{c} \tag{3.8}$$

$$\text{where} \begin{cases} \mathbf{M} &= \left[\sum_{n=1}^{N} \theta_k(y_n)\theta_l(y_n)\right]_{1\le k,l\le K} \\ \mathbf{c} &= \left[\sum_{n=1}^{N} (y_n\theta_k(y_n) - \sigma^2\theta_k'(y_n))\right]_{1\le k\le K} \end{cases}$$

For the remaining of this section, we will drop the location subscript "n" and consequently, y will denote any of the wavelet coefficients y_n.

The difficulty is now to choose suitable basis functions θ_k that will determine the shape of our denoising function. Therefore, we want the denoising function θ to satisfy the following properties:

- **differentiability:** required to apply Corollary 1;
- **anti-symmetry:** the wavelet coefficients are not expected to exhibit a sign preference;
- **linear behavior for large coefficients:** because $\theta(y)$ should asymptotically tend to y.

After trying several types of θ_k, we have found that all of them give quite similar results, when the above conditions are satisfied. We have thus decided to retain the following pointwise denoising function:

$$\theta(y) = \sum_{k=1}^{K} a_k \underbrace{y\exp\left(-(k-1)\frac{y^2}{2T^2}\right)}_{\theta_k(y)} \tag{3.9}$$

We choose derivatives of Gaussians (DOG) because they decay quite fast, which ensures a linear behavior close to the identity for large coefficients (see Figure 3.1).

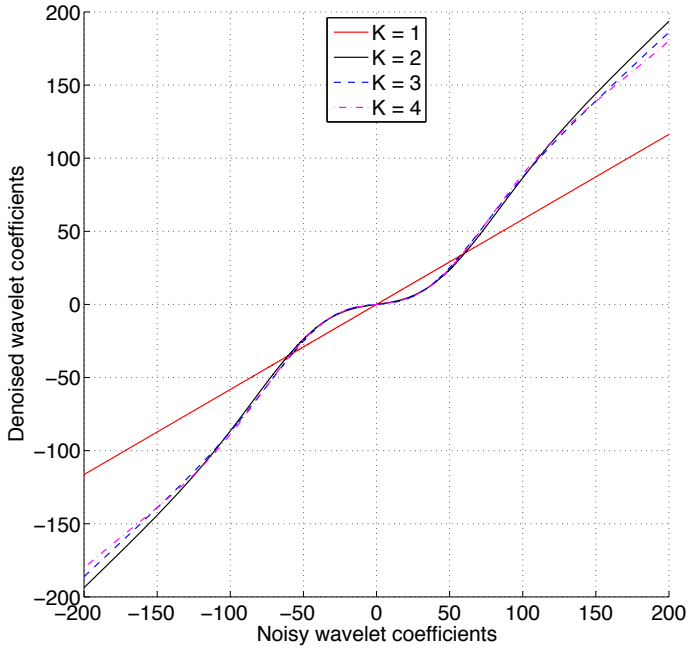


Figure 3.1: The shape of our denoising function (3.9) in a particular wavelet subband, for various K and optimized a_k's and T.

In addition to the linear coefficients, our denoising function contains two nonlinear dependencies: the number of terms K and the parameter T. We will see later that they can be fixed independently of the image.

If we consider only one parameter ($K = 1$), our denoising function simply becomes $\theta(y) = a_1 y$, which is the simplest linear pointwise denoising function. The direct minimization of the estimate ϵ provides

$$\theta(y) = \underbrace{\left(1 - \frac{N\sigma^2}{\|\mathbf{y}\|^2}\right)}_{a_1} y \tag{3.10}$$

The above pointwise linear MMSE estimator can be either interpreted as a pointwise Wiener filter (encountered in Section 1.3.1, Equ. (1.7)) or as a variant of the James-Stein estimator [94].

Practical tests (with optimization over the parameter T, independently in each subband) on various images and with various noise levels have shown that, as soon as

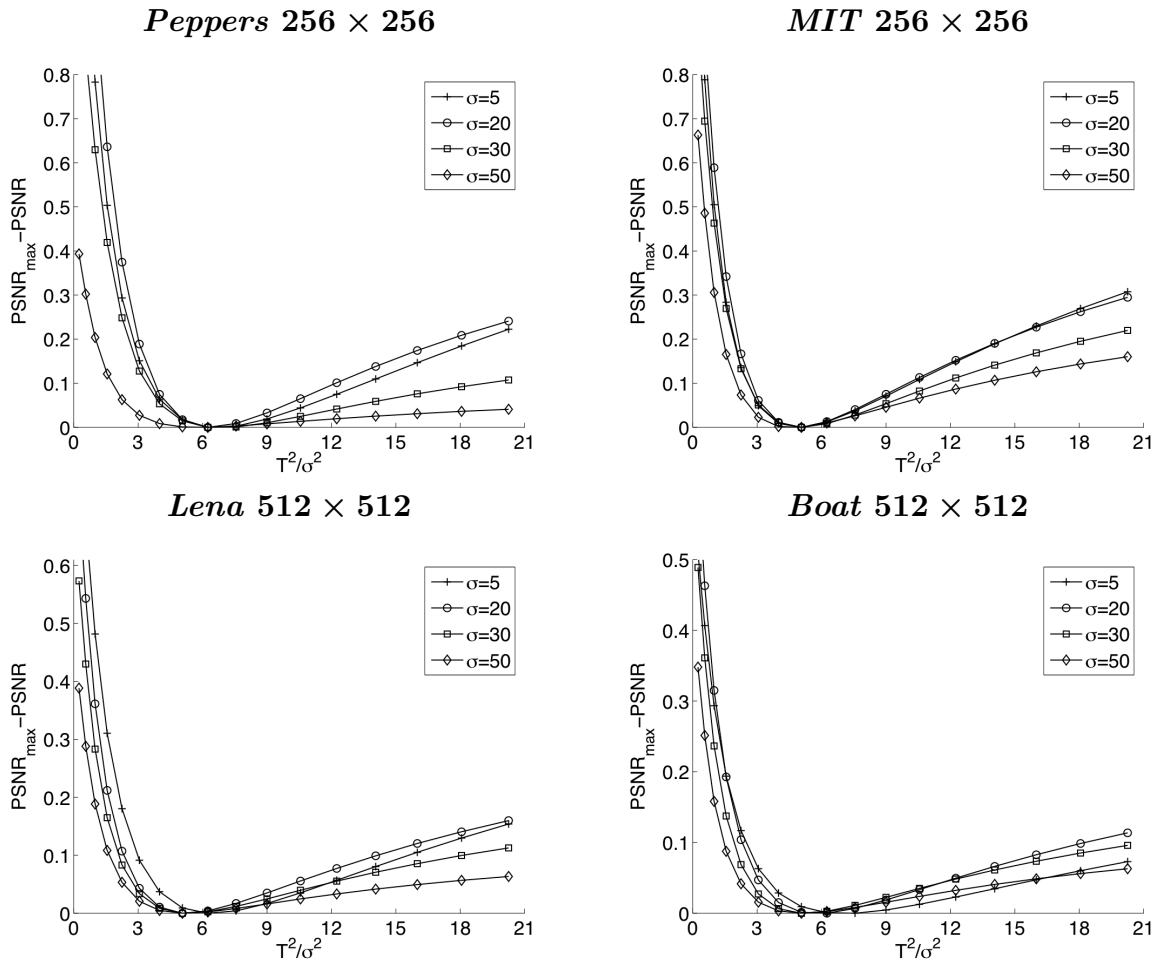


Figure 3.2: Sensitivity of the proposed denoising function (3.11) with respect to variations of T. We can notice that for all images and for the whole range of input PSNR the maximum of the PSNR is reached for $T^2/\sigma^2 \simeq 6$.

$K \geq 2$, the results become quite similar. It thus appears that it is sufficient to keep as few as $K = 2$ terms in (3.9). This is confirmed in Figure 3.1, which shows that the shape of our denoising function is nearly insensitive to the variation of $K \geq 2$.

Moreover, the optimal value of the parameter T is closely linked to the standard deviation σ of the AWGN and in a lesser way to the number of parameters K. Its interpretation is quite similar to the *soft-thresholding* case: it manages the transition between low SNR to high SNR coefficients. In our case though, the variations of the minimal ϵ (over a_k) when T changes are quite small (see Figure 3.2), because our denoising function is much more flexible than the *soft-threshold*. This sensitivity becomes even smaller as the number of parameters K increases. In fact, this indicates that some parameters are in that case useless.

To summarize, we have shown that both the number of terms K and the parameter T have only a minor influence on the quality of the denoising process. This indicates that these two parameters do not have to be optimized; instead, they can be fixed once for all, independently of the type of image. From a practical point of view, we suggest to use $K = 2$ terms and $T = \sqrt{6}\sigma$ (see Figure 3.2), leading to the following pointwise thresholding function:

$$\theta_0(y; \mathbf{a}) = a_1 y + a_2 \exp\left(-\frac{y^2}{12\sigma^2}\right) y \tag{3.11}$$

Now, it is interesting to evaluate the efficiency of our denoising function (3.11) and the accuracy of our minimization process based on a SURE estimate ϵ of the MSE. We propose to compare our results with those obtained by the popular subband-adaptive *soft-threshold* with three different threshold selection rules: **(1)** the threshold is obtained by a general cross validation procedure (*GCVshrink*), as suggested in [95][2]; **(2)** the threshold is obtained by SURE minimization, without restrictions (*SUREshrink**), contrary to the hybrid approach proposed in [71]; **(3)** the threshold is obtained by MSE minimization (*OracleShrink*). Two main observations naturally come out of Table 3.1:

1. SURE is a reliable estimate of the MSE. Indeed, the resulting average loss in PSNR is within 0.1 dB for all images and for a wide range of noise levels. A threshold selection based on SURE minimization is therefore much closer to an *Oracle* optimization than the GCV-based selection.

2. Our sum of DOG (3.11) gives better PSNRs than even the optimal (in the MMSE sense) *soft-threshold*. This indicates that the optimization of two *linear* parameters leads to better results than the optimization of one *non-linear* parameter.

3.3 Interscale Estimator

The integration of interscale information has been shown to improve the denoising quality, both visually and in terms of PSNR [81, 83, 96]. However, the gain brought is often modest, especially considering the additional complications involved by this processing [81]. In this section, we reformulate the problem by first building a loose prediction $\tilde{y}$ of wavelet coefficients y out of a suitably filtered version of the lowpass subband at the same scale, and then by including this predictor in an explicit interscale denoising function. Apart from the specific denoising problem addressed in this thesis, we believe more generally that other applications (e.g. compression, detection, segmentation) could benefit as well from the theory that leads to this predictor.

[2]Matlab implementation available at:
`http://www.cs.kuleuven.be/~maarten/software/pieflab.html`

Table 3.1: Comparison of our sum of DOG (3.11) with three variants of the popular *soft-threshold*.

σ	5	10	20	30	50	100
Method	***Peppers* 256 × 256**					
GCVShrink	34.89	31.42	27.59	25.43	22.97	20.23
SUREShrink*	36.38	32.06	28.02	25.80	23.30	20.69
OracleShrink	*36.39*	*32.08*	*28.04*	*25.83*	*23.34*	*20.78*
Sum of DOG (*SURE*)	**36.67**	**32.37**	**28.27**	**25.95**	**23.43**	**20.94**
Sum of DOG (*Oracle*)	*36.68*	*32.38*	*28.28*	*25.96*	*23.46*	*21.02*
Method	***Bridge* 256 × 256**					
GCVShrink	28.83	26.45	23.97	22.79	21.33	19.42
SUREShrink*	34.83	29.81	25.74	23.89	22.02	19.94
OracleShrink	*34.83*	*29.81*	*25.76*	*23.91*	*22.06*	*20.03*
Sum of DOG (*SURE*)	**34.89**	**30.00**	**26.08**	**24.27**	**22.38**	**20.27**
Sum of DOG (*Oracle*)	*34.89*	*30.00*	*26.09*	*24.28*	*22.40*	*20.34*
Method	***Boat* 512 × 512**					
GCVShrink	34.13	31.19	27.95	26.31	24.36	22.12
SUREShrink*	36.08	32.11	28.64	26.81	24.79	22.50
OracleShrink	*36.09*	*32.12*	*28.65*	*26.82*	*24.81*	*22.55*
Sum of DOG (*SURE*)	**36.35**	**32.38**	**28.86**	**27.03**	**25.02**	**22.75**
Sum of DOG (*Oracle*)	*36.35*	*32.38*	*28.86*	*27.04*	*25.03*	*22.78*
Method	***Goldhill* 512 × 512**					
GCVShrink	32.31	30.05	27.00	25.78	24.34	22.81
SUREShrink*	36.00	31.98	28.75	27.17	25.43	23.38
OracleShrink	*36.00*	*31.98*	*28.76*	*27.19*	*25.46*	*23.44*
Sum of DOG (*SURE*)	**36.22**	**32.25**	**29.00**	**27.43**	**25.68**	**23.67**
Sum of DOG (*Oracle*)	*36.22*	*32.25*	*29.00*	*27.43*	*25.70*	*23.70*

Note: The output PSNRs have been averaged over ten noise realizations.

3.3.1 Building the Interscale Predictor

The wavelet coefficients that lie on the same dyadic tree (see Figure 3.3) are well-known to be take large values in the neighborhood of image discontinuities. What can thus be predicted with reasonably good accuracy are the position of large wavelet coefficients out of parents at lower resolutions. However, getting the actual values of the finer resolution scale coefficients seem somewhat out of reach. This suggests that the best we can get out of interscale correlations is a segmentation between regions of large and small coefficients. This comes back to the idea of signal of interest proposed by Pižurica *et al.* in [82].

In a critically sampled orthonormal wavelet decomposition, the parent subband is half the size of the child subband. The usual way of putting the two subbands in correspondence is simply to expand the parent by a factor two. Unfortunately, this approach does not take into account the potential (non-integer) shift caused by the

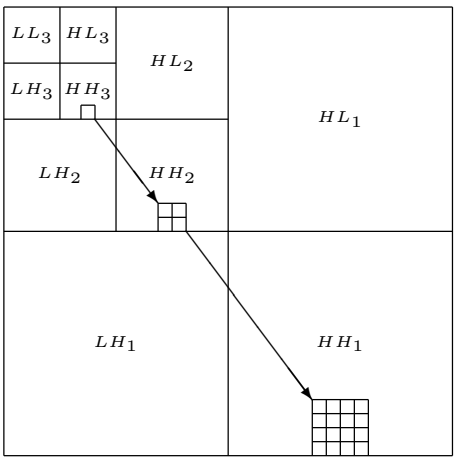


Figure 3.3: Schematic view of three stages of a fully decimated 2D wavelet transform and the so-called parent-child relationship.

filters of the OWT. We thus propose a better solution, which addresses this issue and ensures the alignment of image features between the child and its parent.

Our idea comes from the following observation: let LH_j and LL_j be, respectively, bandpass and lowpass outputs at the decomposition level j of the filterbank. Then, if the *group delay*[3] between the bandpass and the lowpass filters are *equal*, no shift between the features of LH_j and LL_j will occur. Of course, depending on the amplitude response of the filters, some features may be attenuated, blurred, or enhanced, but their relative location will be the same. When the group delays differ, which is the general case, we thus propose to filter the lowpass subband LL_j in order to *compensate for the group delay difference* with LH_j. This operation is depicted in Figure 3.4(A): LL_j is filtered in the three bandpass "directions" by adequately designed filters W_{HL}, W_{HH} and W_{LH}, providing aligned (i.e. group delay compensated) subbands with HL_j, HH_j and LH_j.

Because the filters used in the standard 2D orthonormal wavelet transform are usually separable, we only consider here 1D group delay compensation (GDC).

Definition 1. *We say that two filters $H(z)$ and $G(z)$ are group delay compensated if and only if the group delay of the quotient filter $H(z)/G(z)$ is zero identically, i.e. if and only if there exists a (anti-)symmetric filter $R(z) = \pm R(z^{-1})$ such that $H(z) = G(z)R(z)$.*

The following result shows how to choose a GDC filter in a standard orthonormal filterbank.

[3]The group delay of a filter h with frequency response $H(\boldsymbol{\omega}) = A(\boldsymbol{\omega})\exp(j\phi(\boldsymbol{\omega}))$ is defined as the frequency gradient of its phase response, with a minus sign, i.e. $-\nabla\phi(\boldsymbol{\omega})$.

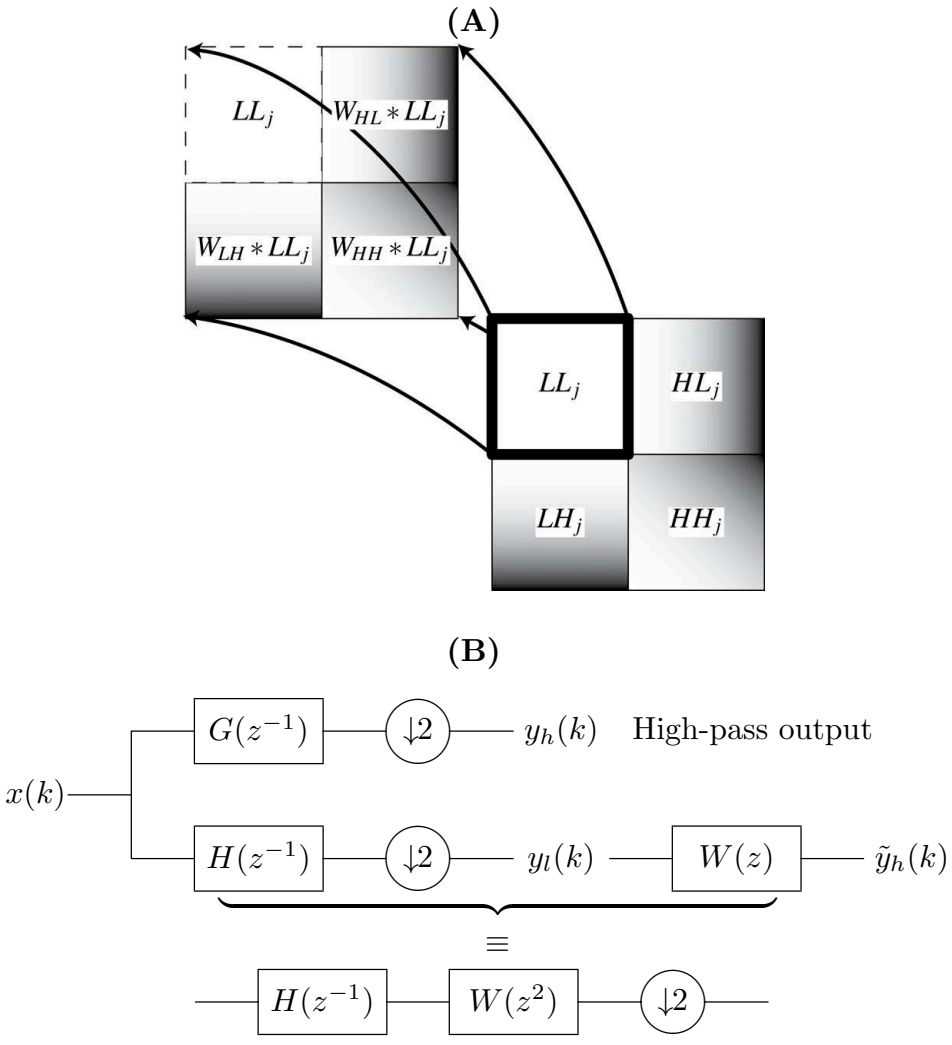


Figure 3.4: One way of obtaining the whole parent information out of the lowpass subband: **(A)** 2D illustration. **(B)** 1D filterbank illustration.

Theorem 3. *For the output of the dyadic orthonormal filterbank of Figure 3.4(B) to be group delay compensated, it is necessary and sufficient that:*

$$W(z^2) = G(z^{-1})G(-z^{-1})(1 + \varepsilon z^{-2})R(z^2) \tag{3.12}$$

where $\varepsilon = \pm 1$ *and* $R(z) = R(z^{-1})$ *is arbitrary.*

Proof. Group delay compensation between the two filterbank branches is equivalent to (see Figure 3.4(B))

$$H(z^{-1})W(z^2) = G(z^{-1})R_1(z) \tag{3.13}$$

where $R_1(z) = \epsilon R_1(z^{-1})$ is an arbitrary symmetric ($\varepsilon = 1$) or anti-symmetric ($\varepsilon = -1$) filter.

Because the filters H and G are orthonormal, we have $H(z^{-1}) = zG(-z)$, and thus (3.13) can be rearranged as:

$$W(z^2) = \frac{G(z^{-1})R_1(z)}{zG(-z)} = G(z^{-1})G(-z^{-1})\frac{z^{-1}R_1(z)}{G(-z)G(-z^{-1})} \tag{3.14}$$

Since both $G(z^{-1})G(-z^{-1})$ and $W(z^2)$ are even polynomial, we can define

$$R_2(z^2) = \frac{z^{-1}R_1(z)}{G(-z)G(-z^{-1})},$$

because the r.h.s. of Equ. (3.14) has to be an even polynomial too.

Using the symmetry of $R_1(z)$, we further observe that $R_2(z) = \varepsilon z^{-1}R_2(z^{-1})$, and consequently, $R_2(z)$ can be factorized as $R_2(z) = (1 + \varepsilon z^{-1})R(z)$, where

$$\begin{aligned} R(z^{-2}) &= \frac{zR_1(z^{-1})}{(1+\varepsilon z^2)G(-z)G(-z^{-1})} \\ &= \frac{\varepsilon z R_1(z)}{(1+\varepsilon z^2)G(-z)G(-z^{-1})} \\ &= \frac{z^{-1}R_1(z)}{(1+\varepsilon z^{-2})G(-z)G(-z^{-1})} \\ &= R(z^2) \end{aligned}$$

i.e. $R(z)$ is an arbitrary zero-phase filter.

After substitution in (3.14), this finally leads us to the formulation (3.12), as an equivalent characterization of the group delay compensation in the filterbank of Figure 3.4(B). ■

Note that the construction of the interscale predictors can be performed by a decimated filterbank similar to that of the standard discrete wavelet transform, where the usual analysis filters $H(z^{-1})$ and $G(z^{-1})$ are replaced by $H(z^{-1})$ and $H(z^{-1})W(z^2)$, respectively.

In addition to (3.12), the GDC filter $W(z)$ has to satisfy a few constraints:

- **Energy preservation**, i.e. $\sum_{n\in\mathbb{Z}} w_n^2 = 1$, in order for the amplitude of the two outputs to be comparable;
- **Highpass behavior**, in order for the filtered lowpass image to "look like" the bandpass target;
- **Shortest possible response**, in order to minimize the enlargement of image features.

We can give a simple GDC filter in the case of symmetric filters. The shortest highpass $W(z)$ satisfying the GDC condition is in fact the simple gradient filter: $W(z) = z - 1$. If the symmetry is not centered at the origin but at a position n_0, then $W(z) = z^{-n_0}(z-1)$. This type of solution is still adequate for near-symmetric filters such as the Daubechies *symlets* [39]. When the lowpass filter is not symmetric, we can simply take $R(z^2) = 1$ in (3.12).

Finally, in order to increase the homogeneity inside regions of similar magnitude coefficients, we apply a 2D-smoothing filter (e.g. a normalized Gaussian kernel $G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$) onto the absolute value of the GDC output. In the rest of this chapter, we will refer to the so-built interscale predictor as $\tilde{y}$ (see Figure 3.5).

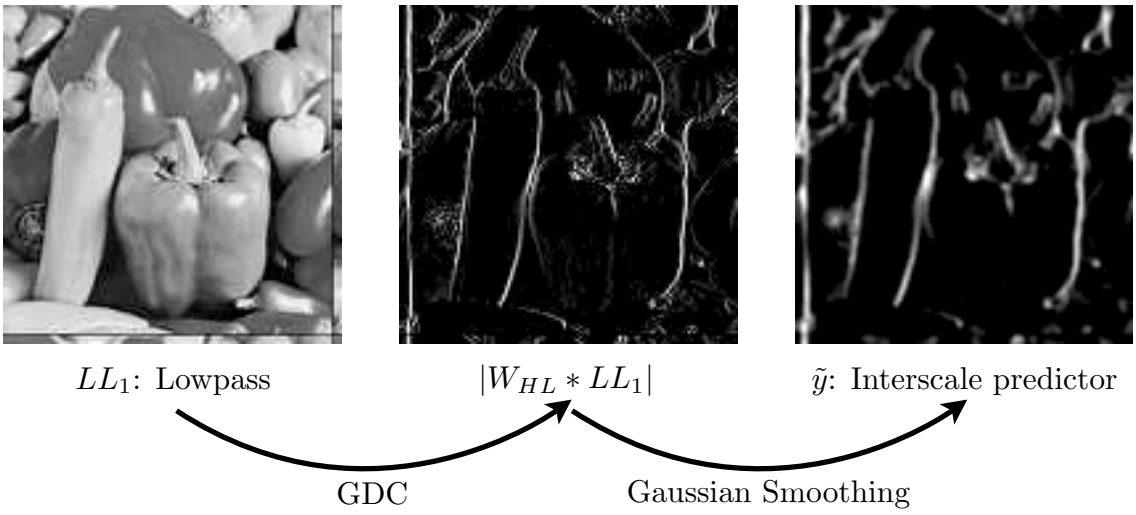


Figure 3.5: Building an efficient interscale predictor, illustrated with a particular subband (HL_1) of the noise-free *Peppers* image.

3.3.2 Integrating the Interscale Predictor

Now that we have built the interscale predictor $\tilde{y}$, we have to suitably integrate it into our pointwise denoising function. As mentioned before, this interscale predictor does not tell us much about the actual value of its corresponding child wavelet coefficients. It only gives an indication on its expected magnitude. Here, we thus propose to use the parent $\tilde{y}$ as a discriminator between high SNR wavelet coefficients and low SNR wavelet coefficients, leading to the following general interscale denoising function:

$$\theta(y, \tilde{y}; \mathbf{a}, \mathbf{b}) = f(\tilde{y}) \sum_{k=1}^{K} a_k \theta_k(y) + (1 - f(\tilde{y})) \sum_{k=1}^{K} b_k \theta_k(y) \tag{3.15}$$

The linear set of parameters **a** and **b** are then solved for by minimizing the MSE estimate ϵ defined in Corollary 1. The optimal coefficients are obtained in the same way as in Section 3.2.2 (i.e. by solving a linear system of $2K$ equations) and involve a solution similar to (3.8).

A first thought choice for the function f in (3.15) is simply the Heaviside function

$$H(\tilde{y}) = \begin{cases} 1, \text{ if } |\tilde{y}| \geq T \\ 0, \text{ if } |\tilde{y}| < T \end{cases} \tag{3.16}$$

where T can be interpreted as a decision factor. However, since the classification will not be perfect (i.e. some small parent coefficients may correspond to high magnitude child coefficients, and vice-versa), it is more appropriate to use a smoother decision function. We thus propose to use instead:

$$f(\tilde{y}) = \exp\left(-\frac{\tilde{y}^2}{2T^2}\right) \tag{3.17}$$

As in the univariate case of Section 3.2.2, we suggest to use a sum of DOG with $K = 2$ terms for each class of wavelet coefficients and [4] $T = \sqrt{6}\sigma$, leading to the following interscale denoising function:

$$\begin{aligned} \theta(y, \tilde{y}; \mathbf{a}, \mathbf{b}) &= \exp\left(-\frac{\tilde{y}^2}{12\sigma^2}\right)\theta_0(y; \mathbf{a}) + \left(1 - \exp\left(-\frac{\tilde{y}^2}{12\sigma^2}\right)\right)\theta_0(y; \mathbf{b}) \\ &= \exp\left(-\frac{\tilde{y}^2}{12\sigma^2}\right)\left(a_1 + a_2 \exp\left(-\frac{y^2}{12\sigma^2}\right)\right) y + \\ &\quad \left(1 - \exp\left(-\frac{\tilde{y}^2}{12\sigma^2}\right)\right)\left(b_1 + b_2 \exp\left(-\frac{y^2}{12\sigma^2}\right)\right) y \end{aligned} \tag{3.18}$$

Table 3.2 quantifies the improvement introduced by this new way of considering the interscale information, as compared to the usual expansion by two of the parent subband.

3.4 Multichannel/Multivariate Estimator

In this section, we propose a vector/matrix extension of SURE-LET denoising in the orthonormal wavelet domain. We now consider N-pixel images with C channels; typically, $C = 3$ color channels for RGB images, but for biological images (fluorescence) or multiband satellite images, C might be much larger. We denote these multichannel images by a $C \times N$ matrix whose columns are the channel values of each pixel:

[4] Side investigations have shown that the T needed in (3.17) and the one optimized in Section 3.2.2 can be chosen identical for optimal performances and equal to $\sqrt{6}\sigma$.

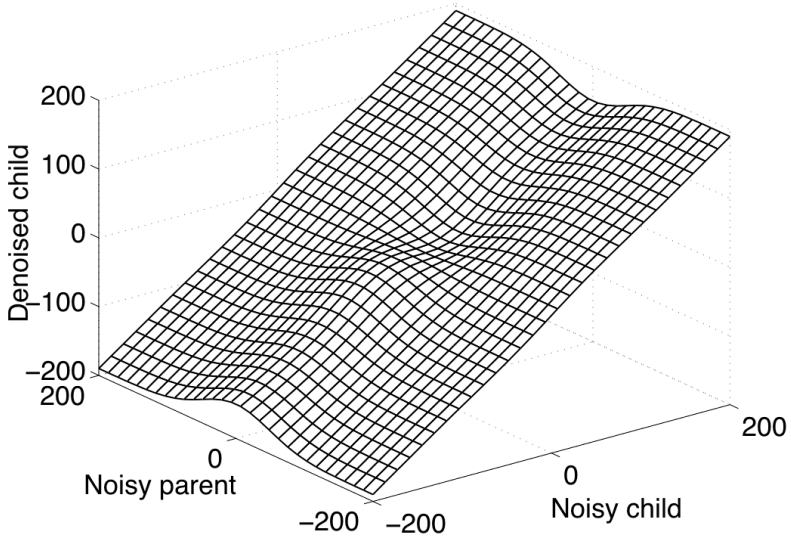


Figure 3.6: 3D surface plot of a possible realization of our interscale thresholding function (3.18).

Table 3.2: Denoising performance improvement brought by our interscale strategy (OWT *sym8*).

$\boldsymbol{\sigma}$	5	10	20	30	50	100
Method	***Peppers* 256 × 256**					
Expansion by 2	36.76	32.49	28.46	26.21	23.62	20.92
Proposed	**37.17**	**33.18**	**29.33**	**27.13**	**24.43**	**21.32**
Method	***House* 256 × 256**					
Expansion by 2	37.50	33.59	30.03	28.07	25.78	22.92
Proposed	**37.88**	**34.29**	**30.93**	**28.98**	**26.58**	**23.51**

Note: The output PSNRs have been averaged over ten noise realizations.

$$\mathbf{x} = [\mathbf{x}_1\ \mathbf{x}_2 \ldots \mathbf{x}_N], \text{ where } \mathbf{x}_n = [x_{1,n}\ x_{2,n}\ \ldots x_{C,n}]^{\mathrm{T}} \tag{3.19}$$

These images are corrupted by an additive channel-wise white Gaussian noise[5] $\mathbf{b} = [\mathbf{b}_1\ \mathbf{b}_2 \ldots \mathbf{b}_N]$ of known $C \times C$ interchannel covariance matrix $\boldsymbol{\Gamma}$, i.e.

$$\mathscr{E}\left\{\mathbf{b}_n \mathbf{b}_{n'}^{\mathrm{T}}\right\} = \boldsymbol{\Gamma}\ \delta_{n-n'}$$

[5] "Channel-wise" means here that the noise is Gaussian and white inside each channel.

We denote the resulting noisy image by $\mathbf{y} = [\mathbf{y}_1\ \mathbf{y}_2 \dots \mathbf{y}_N]$ and we have:

$$\mathbf{y} = \mathbf{x} + \mathbf{b} \tag{3.20}$$

For multichannel images, the MSE between an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ and the actual value of $\mathbf{x}$ can be expressed using Frobenius matrix norm as

$$\begin{aligned} \text{MSE} &= \frac{1}{CN}\|\hat{\mathbf{x}} - \mathbf{x}\|_F^2 \\ &= \frac{1}{CN}\text{trace}\left\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^{\mathrm{T}}\right\} \\ &= \frac{1}{CN}\sum_{n=1}^{N}\|\hat{\mathbf{x}}_n - \mathbf{x}_n\|^2 \end{aligned}$$

After the application of an orthonormal wavelet transform (OWT) to each channel (see Figure 3.7), the resulting multichannel wavelet subbands are denoted by

$$\mathbf{y}^j = \mathbf{x}^j + \mathbf{b}^j,\ j \in [1, J+1] \tag{3.21}$$

The two important conservation properties of the OWT that we recalled at the beginning of this chapter, have the following consequences:

- The wavelet coefficients of the AWGN are Gaussian as well, and are independent within and between the subbands. Moreover, the interchannel covariance matrix remains unchanged

$$\mathscr{E}\left\{\mathbf{b}_n^j {\mathbf{b}_{n'}^{j'}}^{\mathrm{T}}\right\} = \mathbf{\Gamma}\ \delta_{n-n'}\ \delta_{j-j'} \tag{3.22}$$

- The image domain MSE and the subband MSEs are related through

$$\underbrace{\|\hat{\mathbf{x}} - \mathbf{x}\|_F^2}_{CN\times\text{MSE}} = \sum_{j=1}^{J+1}\underbrace{\|\hat{\mathbf{x}}^j - \mathbf{x}^j\|_F^2}_{CN^j\times\text{MSE}^j} \tag{3.23}$$

These two key properties make it particularly attractive to perform independent processing $\boldsymbol{\theta}^j$ in each individual noisy wavelet subbands $\mathbf{y}^j$. To take advantage of both the interchannels similarities and the interscale consistencies that may be intrinsic to $\mathbf{x}$, the thresholding function $\boldsymbol{\theta}^j$ will also involve a multichannel interscale predictor $\tilde{\mathbf{y}}^j$ built out of the lowpass subband at scale j as detailed in Section 3.3.1. We will however remain "pointwise" in the sense that the estimate $\hat{\mathbf{x}}_n^j$ of the n^{th} pixel of subband j will depend only on $\mathbf{y}_n^j$ and $\tilde{\mathbf{y}}_n^j$, without taking their neighbours into account. It is essential to recall that, because of the statistical independence between the noise component from wavelet subbands of different iteration depth, $\mathbf{y}^j$ and $\tilde{\mathbf{y}}^j$ are also statistically independent.

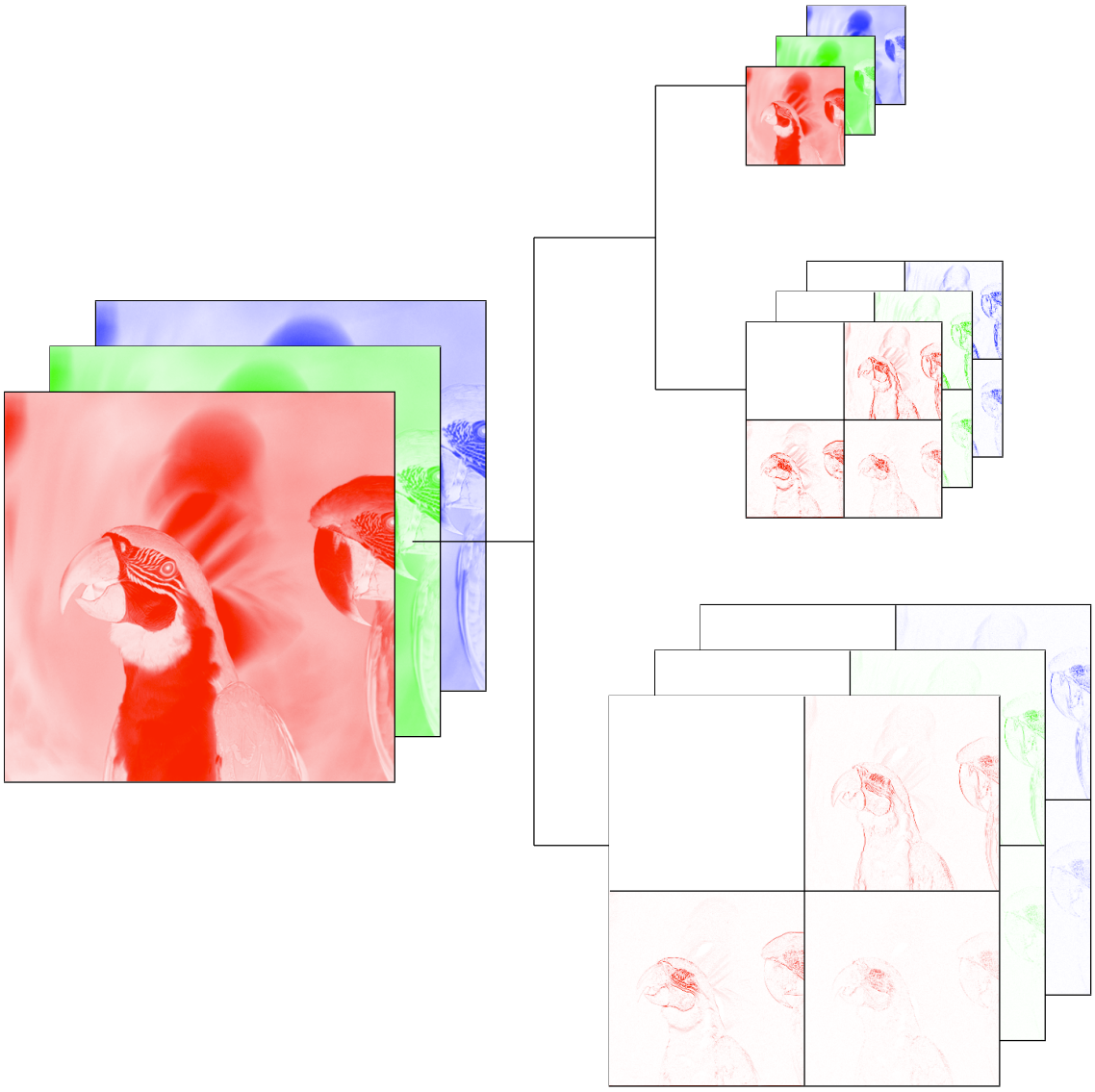


Figure 3.7: Two iterations of a 2-D orthonormal wavelet transform applied to a RGB image.

From now on, we will drop the subband superscript j when no ambiguity is likely to arise. More abstractly, we are thus going to consider the denoising of a multichannel (subband) image $\mathbf{y} = \mathbf{x} + \mathbf{b}$, given an independent prediction (parent) $\tilde{\mathbf{y}}$, by using a $\mathbb{R}^C \times \mathbb{R}^C \to \mathbb{R}^C$ function $\boldsymbol{\theta}$ relating the coefficients of $\mathbf{y}$ and $\tilde{\mathbf{y}}$ to the coefficients of the estimate $\hat{\mathbf{x}}$ through

$$\hat{\mathbf{x}}_n = \boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n), \text{ for } n = 1, 2, \ldots, N \tag{3.24}$$

3.4.1 Multichannel SURE

In this section, we derive an unbiased MSE estimate for a multichannel interscale thresholding applied in the orthonormal wavelet domain.

Corollary 2. *Assume that $\boldsymbol{\theta}(\cdot,\cdot)$ is (weakly) differentiable w.r.t. its first variable, and such that $\|\boldsymbol{\theta}(\mathbf{u},\mathbf{v})\| \leq cste(\mathbf{v}) \times \exp(\|\mathbf{u}\|^2/(2s^2))$ where $s > \sigma$. Then, if the estimate $\hat{\mathbf{x}}$ is built according to* (3.24)*, the following random variable*

$$\begin{aligned} \epsilon &= \frac{1}{CN}\sum_{n=1}^{N}\|\boldsymbol{\theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)-\mathbf{y}_n\|^2 + \\ &\quad \frac{2}{CN}\sum_{n=1}^{N}\mathrm{trace}\left\{\boldsymbol{\Gamma}^{\mathrm{T}}\nabla_1\boldsymbol{\theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)\right\} - \frac{1}{C}\mathrm{trace}\left\{\boldsymbol{\Gamma}\right\} \end{aligned} \tag{3.25}$$

is an unbiased estimator of the expected MSE, i.e.

$$\mathscr{E}\left\{\epsilon\right\} = \frac{1}{CN}\mathscr{E}\left\{\|\hat{\mathbf{x}}-\mathbf{x}\|_F^2\right\}$$

Here, we have denoted by $\nabla_1\boldsymbol{\theta}$ the matrix containing the partial derivatives of the components of $\boldsymbol{\theta} = [\theta_1\ \theta_2 \ldots \theta_C]^{\mathrm{T}}$ with respect to its *first* variable:

$$\nabla_1\boldsymbol{\theta}(\mathbf{u},\mathbf{v}) = \begin{bmatrix} \frac{\partial\theta_1(\mathbf{u},\mathbf{v})}{\partial u_1} & \frac{\partial\theta_2(\mathbf{u},\mathbf{v})}{\partial u_1} & \cdots & \frac{\partial\theta_C(\mathbf{u},\mathbf{v})}{\partial u_1} \\ \frac{\partial\theta_1(\mathbf{u},\mathbf{v})}{\partial u_2} & \frac{\partial\theta_2(\mathbf{u},\mathbf{v})}{\partial u_2} & \cdots & \frac{\partial\theta_C(\mathbf{u},\mathbf{v})}{\partial u_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial\theta_1(\mathbf{u},\mathbf{v})}{\partial u_C} & \frac{\partial\theta_2(\mathbf{u},\mathbf{v})}{\partial u_C} & \cdots & \frac{\partial\theta_C(\mathbf{u},\mathbf{v})}{\partial u_C} \end{bmatrix}$$

Proof. Note that, because the noise contaminating $\tilde{\mathbf{y}}_n$ is independent from the one degrading $\mathbf{y}_n$, we may simply prove the result without considering $\tilde{\mathbf{y}}_n$ to be random. We can then develop the squared error between $\mathbf{x}_n = \mathbf{y}_n - \mathbf{b}_n$ and its estimate $\hat{\mathbf{x}}_n = \boldsymbol{\theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)$ as

$$\begin{aligned} \mathscr{E}\left\{\|\boldsymbol{\theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)-\mathbf{x}_n\|^2\right\} &= \mathscr{E}\left\{\|\boldsymbol{\theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)\|^2\right\} \\ &\quad -2\mathscr{E}\left\{\boldsymbol{\theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)^{\mathrm{T}}(\mathbf{y}_n-\mathbf{b}_n)\right\} + \|\mathbf{x}_n\|^2 \\ &= \mathscr{E}\left\{\|\boldsymbol{\theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)-\mathbf{y}_n\|^2\right\} \\ &\quad +2\mathscr{E}\left\{\boldsymbol{\theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)^{\mathrm{T}}\mathbf{b}_n\right\} + \|\mathbf{x}_n\|^2 - \mathscr{E}\left\{\|\mathbf{y}_n\|^2\right\} \end{aligned} \tag{3.26}$$

Now we use the fact that a zero-mean multivariate Gaussian probability density function $q(\mathbf{b}_n)$ with covariance matrix $\boldsymbol{\Gamma}$ satisfies $\mathbf{b}_n q(\mathbf{b}_n) = -\boldsymbol{\Gamma}\nabla q(\mathbf{b}_n)$ to evaluate

$\mathscr{E}\left\{\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)^{\mathrm{T}}\mathbf{b}_n\right\}$:

$$\begin{aligned}
&\mathscr{E}\left\{\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)^{\mathrm{T}}\mathbf{b}_n\right\} \\
&= \int_{\mathbb{R}^C} \boldsymbol{\theta}(\mathbf{x}_n + \mathbf{b}_n, \tilde{\mathbf{y}}_n)^{\mathrm{T}}\mathbf{b}_n q(\mathbf{b}_n)\,\mathrm{d}\mathbf{b}_n \\
&= -\int_{\mathbb{R}^C} \boldsymbol{\theta}(\mathbf{x}_n + \mathbf{b}_n, \tilde{\mathbf{y}}_n)^{\mathrm{T}}\boldsymbol{\Gamma}\,\nabla q(\mathbf{b}_n)\,\mathrm{d}\mathbf{b}_n \\
&= \int_{\mathbb{R}^C} \mathrm{div}_{\mathbf{b}_n}\Big\{\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{x}_n + \mathbf{b}_n, \tilde{\mathbf{y}}_n)\Big\}\, q(\mathbf{b}_n)\,\mathrm{d}\mathbf{b}_n \;\text{ (by parts)} \\
&= \int_{\mathbb{R}^C} \mathrm{trace}\left\{\boldsymbol{\Gamma}^{\mathrm{T}}\nabla_1\boldsymbol{\theta}(\mathbf{x}_n + \mathbf{b}_n, \tilde{\mathbf{y}}_n)\right\} q(\mathbf{b}_n)\,\mathrm{d}\mathbf{b}_n \\
&= \mathscr{E}\left\{\mathrm{trace}\left\{\boldsymbol{\Gamma}^{\mathrm{T}}\nabla_1\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\right\}\right\}
\end{aligned}$$

Introducing the above relation, as well as the standard result

$$\mathscr{E}\left\{\|\mathbf{y}_n\|^2\right\} \;=\; \|\mathbf{x}_n\|^2 + \mathrm{trace}\left\{\boldsymbol{\Gamma}\right\}$$

into (3.26), leads us to the desired result. ■

3.4.2 Multichannel SURE-LET

In this section, we show how to adapt the monochannel SURE-LET denoiser introduced in Section 3.3 to multichannel image denoising. The two fundamental ingredients of the SURE-LET approach remain the same:

1. The denoising function $\boldsymbol{\theta}$ is built as a *linear expansion* of simple (possibly non-linear) thresholding functions $\boldsymbol{\theta}_k$:

$$\begin{aligned}
\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n) &= \sum_{k=1}^{K} \mathbf{a}_k^{\mathrm{T}}\boldsymbol{\theta}_k(\mathbf{y}_n, \tilde{\mathbf{y}}_n) \\
&= \underbrace{[\mathbf{a}_1^{\mathrm{T}}\ \mathbf{a}_2^{\mathrm{T}}\ \ldots \mathbf{a}_K^{\mathrm{T}}]}_{\mathbf{A}^{\mathrm{T}}} \times \underbrace{\begin{bmatrix} \boldsymbol{\theta}_1(\mathbf{y}_n, \tilde{\mathbf{y}}_n) \\ \boldsymbol{\theta}_2(\mathbf{y}_n, \tilde{\mathbf{y}}_n) \\ \vdots \\ \boldsymbol{\theta}_K(\mathbf{y}_n, \tilde{\mathbf{y}}_n) \end{bmatrix}}_{\boldsymbol{\Theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)}
\end{aligned} \tag{3.27}$$

   Here, $\boldsymbol{\Theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)$ is a $KC \times 1$ vector, the $\mathbf{a}_k$'s are $C \times C$ matrices and hence $\mathbf{A}$ is a $KC \times C$ matrix. In this formalism the gradient of $\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)$ with respect to the first variable can be expressed as

$$\nabla_1\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n) = \nabla_1\boldsymbol{\Theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\,\mathbf{A}$$

2. The MSE estimate ϵ is *quadratic* in $\mathbf{A}$, as shown below:

$$\begin{aligned}
\epsilon &= \frac{1}{CN}\sum_{n=1}^{N}\left\|\mathbf{A}^{\mathrm{T}}\boldsymbol{\Theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)-\mathbf{y}_n\right\|^2 \\
&\quad +\frac{2}{CN}\sum_{n=1}^{N}\mathrm{trace}\left\{\boldsymbol{\Gamma}^{\mathrm{T}}\nabla_1\boldsymbol{\Theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)\,\mathbf{A}\right\}-\frac{1}{C}\mathrm{trace}\left\{\boldsymbol{\Gamma}\right\} \\
&= \frac{1}{CN}\sum_{n=1}^{N}\mathrm{trace}\left\{\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)-\mathbf{y}_n\right)\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)-\mathbf{y}_n\right)^{\mathrm{T}}\right\} \\
&\quad +\frac{2}{CN}\sum_{n=1}^{N}\mathrm{trace}\left\{\boldsymbol{\Gamma}^{\mathrm{T}}\nabla_1\boldsymbol{\Theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)\,\mathbf{A}\right\}-\frac{1}{C}\mathrm{trace}\left\{\boldsymbol{\Gamma}\right\} \\
&= \frac{1}{CN}\mathrm{trace}\left\{\mathbf{A}^{\mathrm{T}}\mathbf{M}\mathbf{A}-2\mathbf{C}^{\mathrm{T}}\mathbf{A}\right\}+\frac{1}{CN}\mathrm{trace}\left\{\mathbf{y}\mathbf{y}^{\mathrm{T}}-N\boldsymbol{\Gamma}\right\} \qquad (3.28)
\end{aligned}$$

where we have defined

$$\left\{\begin{array}{rcl}
\mathbf{M} &=& \displaystyle\sum_{n=1}^{N}\boldsymbol{\Theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)\boldsymbol{\Theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)^{\mathrm{T}} \\
\mathbf{C} &=& \displaystyle\sum_{n=1}^{N}\Big(\boldsymbol{\Theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)\mathbf{y}_n^{\mathrm{T}}-\big(\nabla_1\boldsymbol{\Theta}(\mathbf{y}_n,\tilde{\mathbf{y}}_n)\big)^{\mathrm{T}}\boldsymbol{\Gamma}\Big)
\end{array}\right. \qquad (3.29)$$

Finally, the minimization of (3.28) with respect to $\mathbf{A}$ boils down to the following linear system of equations:

$$\mathbf{A}_{\mathrm{opt}}=\mathbf{M}^{-1}\mathbf{C} \qquad (3.30)$$

Obviously, it is advisable to have the smallest possible number of parameters, in order for $\mathbf{M}$ to be nonsingular and in order to maintain the variance of ϵ as small as possible, in such a way that any of its realizations is close to the actual MSE.

It may also be interesting to restrict the number of degrees of freedom of the coefficient matrices $\mathbf{a}_k$ and, in exchange, increase the actual number K of these coefficients: typically, one may choose $\mathbf{a}_k$ to be of the form $\mathbf{u}_k\mathbf{v}_k^{\mathrm{T}}$ where $\mathbf{v}_k$ is some *known* $C\times 1$ vector, while $\mathbf{u}_k$ is an *unknown* $C\times 1$ vector. This means that the $KC\times C$ matrix $\mathbf{A}$ lives in some linear subspace of dimension $D<KC^2$ spanned by, say, a basis of $KC\times C$ matrices $\{\mathbf{E}_d\}_{d=1,2,\ldots,D}$. Once again, minimizing (3.28) with respect to all the degrees of freedom of $\mathbf{A}$ leads to a linear system of equations

$$\mathrm{trace}\left\{\mathbf{E}_d^{\mathrm{T}}\big(\mathbf{M}\mathbf{A}_{\mathrm{opt}}-\mathbf{C}\big)\right\}=0 \qquad \text{for } d=1,2,\ldots D,$$

from which the D (linear) degrees of freedom of $\mathbf{A}_{\mathrm{opt}}$ can be computed.

3.4.3 Multichannel Interscale Thresholding

We propose now a natural vectorization of the interscale thresholding defined in Equ. (3.18) by taking into account the strong similarities that may occur between the various channels. More specifically, we build this thresholding function according to the expression (3.27) with $K = 4$ in which each $\boldsymbol{\theta}_k$ denoises a particular zone of the multichannel wavelet subband, characterized by large or small values of the parents/wavelet coefficients. This zone selection makes use of a "trigger" function $\gamma(x)$ which is essentially unity for small values of $|x|$, and vanishes for large values. We have chosen the following expression:

$$\gamma(x) = \exp\left(-\frac{|x|}{12\sqrt{C}}\right) \tag{3.31}$$

The interscale predictor $\tilde{\mathbf{y}}$ will then be used in order to smoothly discriminate between high-SNR and low-SNR wavelet coefficients, which finally leads to the following multichannel interscale thresholding function:

$$\begin{aligned}
\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n) \quad = \quad & \underbrace{\gamma(\tilde{\mathbf{y}}_n^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\tilde{\mathbf{y}}_n)\gamma(\mathbf{y}_n^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\mathbf{y}_n)}_{\text{small parents and small coefficients}} \quad \mathbf{a}_1^{\mathrm{T}}\mathbf{y}_n \; + \\
& \underbrace{\big(1-\gamma(\tilde{\mathbf{y}}_n^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\tilde{\mathbf{y}}_n)\big)\gamma(\mathbf{y}_n^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\mathbf{y}_n)}_{\text{large parents and small coefficients}} \quad \mathbf{a}_2^{\mathrm{T}}\mathbf{y}_n \; + \\
& \underbrace{\gamma(\tilde{\mathbf{y}}_n^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\tilde{\mathbf{y}}_n)\big(1-\gamma(\mathbf{y}_n^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\mathbf{y}_n)\big)}_{\text{small parents and large coefficients}} \quad \mathbf{a}_3^{\mathrm{T}}\mathbf{y}_n \; + \\
& \underbrace{\big(1-\gamma(\tilde{\mathbf{y}}_n^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\tilde{\mathbf{y}}_n)\big)\big(1-\gamma(\mathbf{y}_n^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\mathbf{y}_n)\big)}_{\text{large parents and large coefficients}} \quad \mathbf{a}_4^{\mathrm{T}}\mathbf{y}_n
\end{aligned} \tag{3.32}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ and $\mathbf{a}_4$ are $C \times C$ matrices, leading to an overall number of $4C^2$ parameters.

When $C = 1$, we recover the monochannel interscale thresholding (3.18). In the experimental section 3.7.2, we have retained the above expression because of its simplicity. However, we have observed that by increasing K from 4 to $4C$ (by increasing the number of zones, e.g. by distinguishing between parents in the same channel from parents in other channels) and decreasing the number of degrees of freedom of the coefficients $\mathbf{a}_k$ from $C \times C$ full-rank matrices to $C \times C$ matrices having non-zero elements in a single column (the overall number of parameters thus remains $4C^2$) yields often better denoising results that may in some cases reach up to $+0.3\,\mathrm{dB}$.

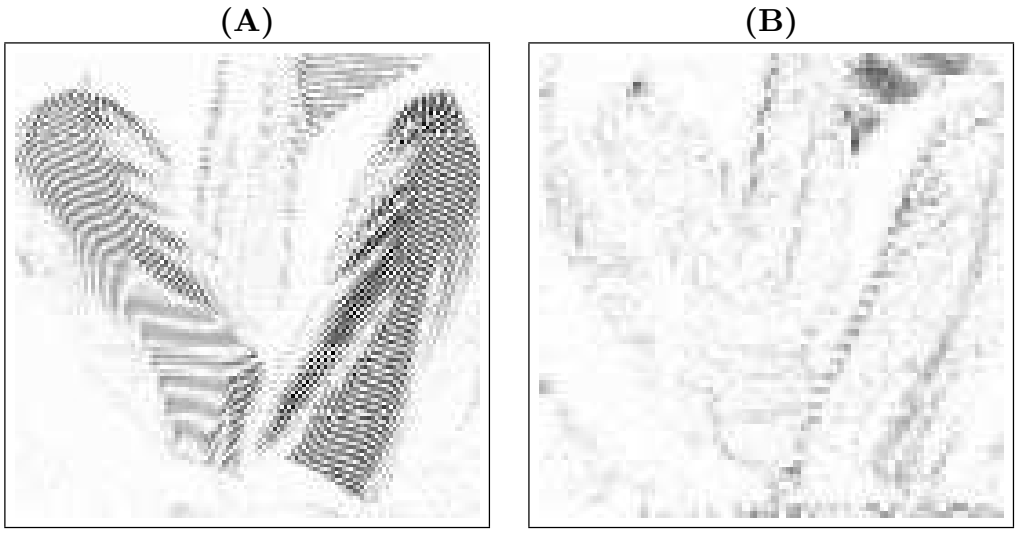


Figure 3.8: **(A)** A zoom at *Barbara*'s trousers at the finest scale of an orthonormal wavelet transform: the stripes are clearly visible. **(B)** A zoom at *Barbara*'s trousers at the next coarser scale (expanded by two): the stripes are not visible anymore.

3.4.4 From Multichannel to Multivariate Interscale Thresholding

In Section 3.3, we have shown that an efficient integration of the interscale dependencies can bring a significant improvement, both quantitatively (higher PSNR) and qualitatively (better preservation of edges). However, some particular image features (such as textures) can have a relatively tight frequency localization, and might thus be only present at a given resolution in a particular bandpass wavelet subband (see Figure 3.8). Interscale dependencies of such patterns might therefore be very weak, if not inexistant. To efficiently denoise grayscale images with such content, the *intrascale* relationships between neighboring pixels must be definitively accounted for.

In practice, a vector $\mathbf{y}_n$ of $C = M^2$ wavelet coefficients taken from a $M \times M$ square neighborhood centred around the location n is formed (see Figure 3.9). In most Bayesian approaches (e.g. [81,84]), these vectors of neighboring coefficients are used to estimate the local covariance matrix $\mathbf{\Gamma_x}$ of the underlying noise-free wavelet coefficients as:

$$\mathbf{\Gamma_x} = \underbrace{\frac{1}{N}\sum_{n=1}^{N} \mathbf{y}_n \mathbf{y}_n^{\mathrm{T}}}_{\mathbf{\Gamma_y}} - \mathbf{\Gamma}$$

where $\mathbf{\Gamma}$ is the covariance matrix of the noise $\mathbf{b}$ as defined in (3.22). Recall that in

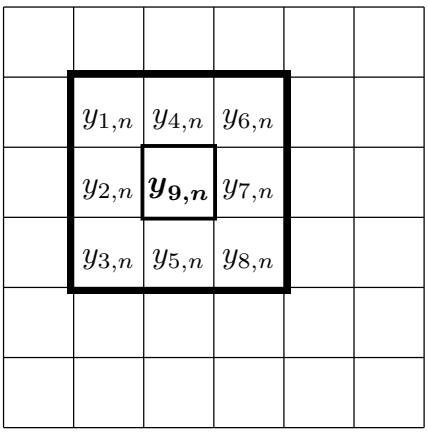


Figure 3.9: Construction of a $C \times 1$ multivariate coefficient $\mathbf{y}_n$ from its $M \times M$ square neighborhood (here, $M = 3$ and thus $C = 9$). The central coefficient $y_{C,n}$ is denoted by $\mathbf{e}_C^{\mathrm{T}}\mathbf{y}_n$.

an orthonormal wavelet representation, each AWGN coefficient is independent from all of its neighbors: hence, $\mathbf{\Gamma} = \sigma^2\mathbf{Id}$ inside each wavelet subband. The estimated covariance matrix of the noise-free wavelet coefficients $\mathbf{\Gamma_x}$ is then often used to derive a multivariate maximum a posteriori [84] or a multivariate Bayesian least-square [81] estimator.

In our approach, we propose to re-interpret *multivariate* denoising as a particular case of *multichannel* denoising, where the $C = M^2$ channels of $\mathbf{y}_n$ are formed by the $C - 1$ neighbors of the central coefficient and the central coefficient $y_{n,C} = \mathbf{e}_C^{\mathrm{T}}\mathbf{y}_n$ itself. The main difference from the multichannel processing that we previously described, is that we only need to estimate the central coefficient, i.e. $\mathbf{e}_C^{\mathrm{T}}\hat{\mathbf{x}}_n = \mathbf{e}_C^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)$. In the multivariate case, the multichannel SURE-LET procedure described in Section 3.4.2 is still valid, except that now

$$\begin{cases} \mathbf{M} &= \displaystyle\sum_{n=1}^{N} \mathbf{E}_C^{\mathrm{T}}\mathbf{\Theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\mathbf{\Theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)^{\mathrm{T}}\mathbf{E}_C \\ \mathbf{C} &= \displaystyle\sum_{n=1}^{N} \mathbf{E}_C^{\mathrm{T}}\Big(\mathbf{\Theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\mathbf{y}_n^{\mathrm{T}} - \big(\nabla_1\mathbf{\Theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\big)^{\mathrm{T}}\mathbf{\Gamma}\Big) \end{cases}$$

where $\mathbf{E}_C = \begin{bmatrix} \mathbf{e}_C & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{e}_C & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{e}_C \end{bmatrix}$ is a $KC \times K$ matrix which restricts the degrees of freedom of $\mathbf{A}_{\mathrm{opt}} = \mathbf{M}^{-1}\mathbf{C}$ from $KC \times C$ to $K \times C$.

Considering the interscale thresholding (3.32) (i.e. $K = 4$) with a standard

square neighborhood of coefficients (e.g. $C = 3 \times 3$) would lead to 36 parameters per wavelet subband. This over-parametrization would significantly increase the variance of SURE, leading to an unreliable parameters optimization. To avoid this scenario, we propose to use cross-shaped neighborhood instead. This choice of shape is further justified by the natural directional selectivity of the separable 2D OWT, which favors vertical and horizontal features. To check if this directional selectivity is reflected in the structure of the optimal linear parameters $\mathbf{A}_{\text{opt}}$, we have considered a 5×5 square neighborhood and optimized, in the MMSE sense, the weights of the linear combination of these 25 coefficients. The magnitudes of these optimal weights $\mathbf{A}_{\text{opt}}$, rearranged as 5×5 matrix, are displayed in Figure 3.10 for various wavelet subbands of two standard grayscale images *Peppers* and *Lena*. We can make the two following observations:

- As expected, the magnitude of the optimal coefficients is higher along the two preferential directions (i.e. either horizontal, vertical or both) and maximum for the central coefficient. A cross-shaped neighborhood thus constitutes an appropriate choice.

- Remarkably, the values of the optimal coefficients are roughly symmetric along the two preferential directions.

The consideration of the above observations allows us to finally reduce the total number of parameters per wavelet subband from $K \times M^2$ to $K \times M$, by imposing the following structure on each of the parameters matrix $\mathbf{a}_k$:

$$\mathbf{a}_k = \begin{bmatrix} 0 & \dots & 0 & a_1^k & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & a_{M_0}^k & 0 & \dots & 0 \\ a_{M_0+1}^k & \dots & a_{M-1}^k & a_M^k & a_{M-1}^k & \dots & a_{M_0+1}^k \\ 0 & \dots & 0 & a_{M_0}^k & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & a_1^k & 0 & \dots & 0 \end{bmatrix}, \tag{3.33}$$

where $M_0 = \dfrac{M-1}{2}$, M odd.

To summarize, for multivariate denoising, we propose to use the multivariate interscale thresholding defined in Equ. (3.32) with a $M \times M$ cross-shaped neighborhood: in practice, we suggest to use $M = 3$ (resp. $M = 5$) for 256×256 (resp. 512×512) images. Inside this neighborhood, only M degrees of freedom are allowed for each parameter $\mathbf{a}_k$ to account for the symmetries observed in Figure 3.10.

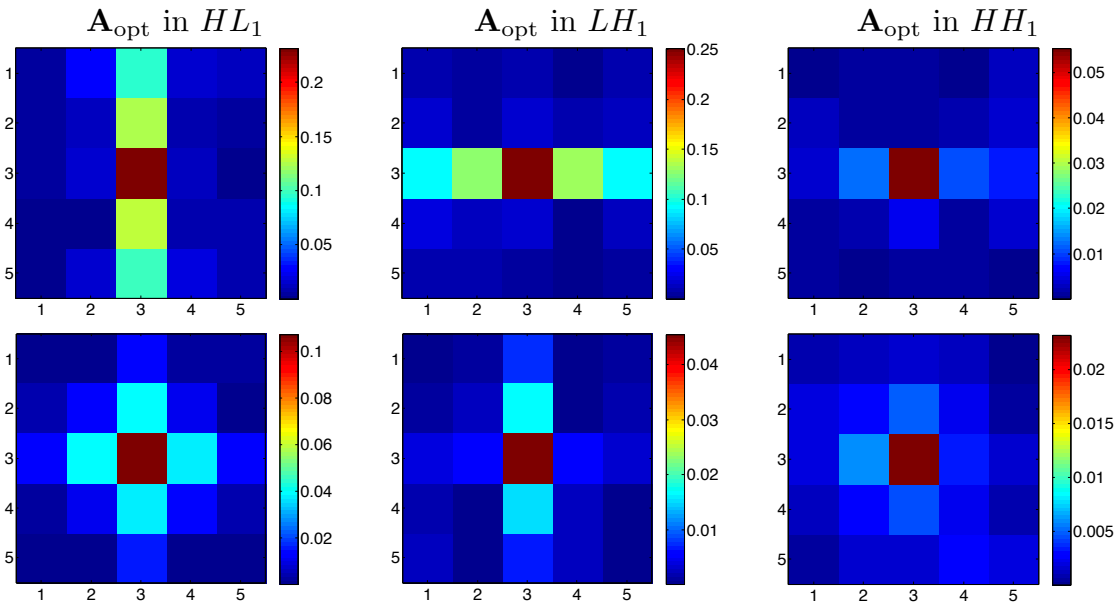


Figure 3.10: Illustration of the directional structure of the optimal (in the MMSE sense) set of parameters $\mathbf{A}_{\text{opt}}$ for the highpass subbands at the finest resolution. Top row: *Peppers*. Bottom row: *Lena*.

3.5 SURE-LET Estimators: a Comparison

Before comparing our SURE-LET denoising strategy with some state-of-the-art algorithms, it is instructive to evaluate the performance of the various grayscale wavelet estimators proposed in this chapter. In Figure 3.11, we compare the results obtained with the simple two-terms sum of DOG (3.11), the interscale thresholding (3.18) and the multivariate interscale estimator (3.32) described in Section 3.4.4. The improvement (often more than +1 dB) brought by the integration of interscale dependencies is quite significant for most standard images. Yet, for images that have a substantial well-localized frequency content (e.g. *Barbara*), the integration of interscale dependencies does not lead to such an impressive gain. As expected for such images, the integration of intrascale dependencies considerably increases the denoising performance (around +1.5 dB). In most other cases, it usually brings an improvement of +0.5 dB over the interscale estimator.

3.6 Noise Variance Estimation

Most of the denoising algorithms designed for AWGN reduction require the knowledge of the AWGN variance σ^2. Numerous works are thus entirely dedicated to this estimation problem (e.g. [97–102]).

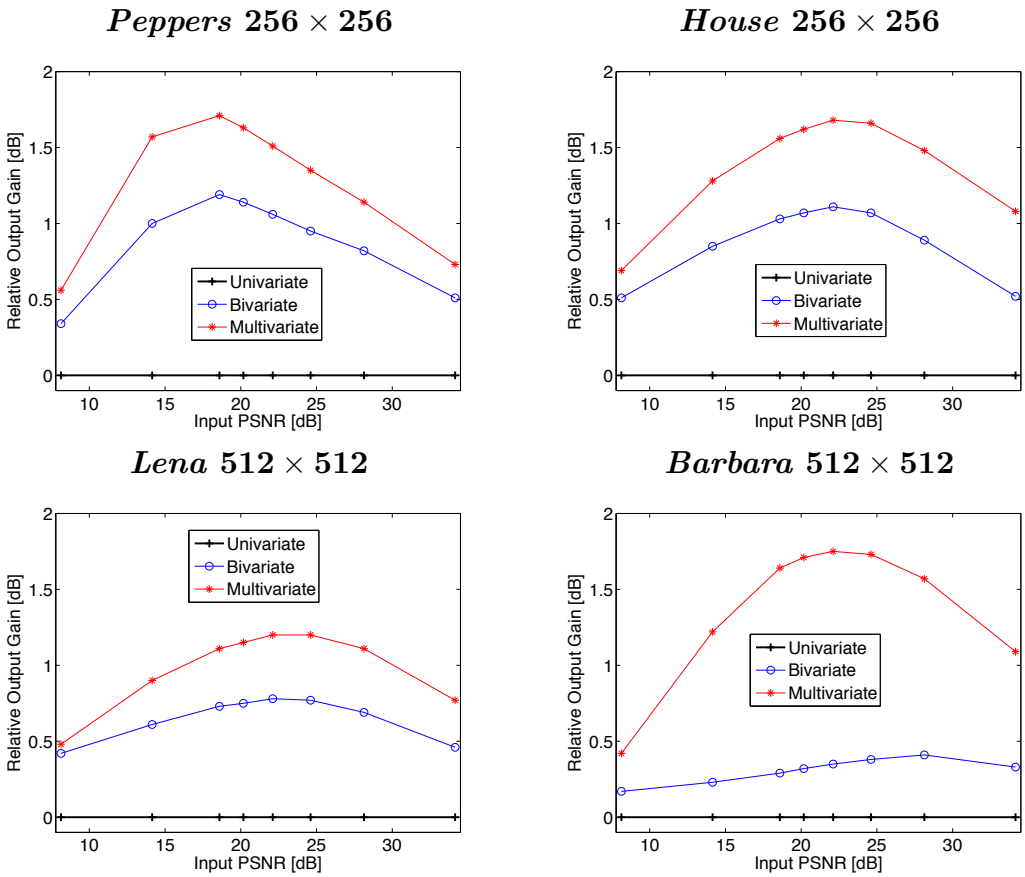


Figure 3.11: Comparison of the various wavelet estimators introduced in this chapter. The benchmark is provided by the simple univariate sum of DOG (3.11) ("+" markers); the interscale thresholding (3.18) is referred to as "bivariate" ("∘" markers) and its multivariate extension (3.32) as "multivariate" ("∗" markers).

Two strategies are usually considered: the first one consists in computing the local variance inside every $M \times M$ (typically $M = 8$) blocks of the noisy image. The AWGN variance is then estimated by averaging a subset of these local variances. The retained subset should only contain the noise variances computed inside the most homogenous regions of the image. The second one consists in prefiltering the noisy image with a unit-norm highpass filter, and then estimating the noise variance from the filtered residual. In this approach, the residual is assumed to mainly contain the AWGN noise component. This estimation method has been popularized in wavelet-based denoising algorithms by Donoho, who proposed the median of the absolute deviation (MAD) of the highest frequency subband (HH) as an estimate $\hat{\sigma}$ of the

noise standard deviation [103]:

$$\hat{\sigma} = 1.4826 \text{ med}\left\{|\mathbf{y} - \text{med}\left\{\mathbf{y}\right\}|\right\},\ y_n \in \text{HH} \tag{3.34}$$

For a wide range of noise levels, the wavelet-domain MAD estimator usually gives an accurate estimation of the AWGN variance for most natural images. However, under relatively low noise conditions ($\sigma < 10$), it becomes unreliable for those images having a significant high-frequency content. Indeed, in that case, the noise variance will be overestimated due to the non-negligeable contribution of the noise-free signal energy present in the highest frequency noisy wavelet subband. To overcome this issue, we propose to design a linear filter that minimizes the energy of the noise-free signal, while still preserving the AWGN variance[6].

Proposition 4. *The unit-norm filter* $\mathbf{h} \in \mathbb{R}^M$ *that minimizes the residual energy of the noise-free signal* $\|\mathbf{h} * \mathbf{x}\|^2$ *is the eigenvector corresponding to the minimum eigenvalue of the symmetric matrix* $\mathbf{A} = [a_{i,j}]_{1 \leq i,j \leq M}$*, where* $a_{i,j} = \sum_{n=1}^{N} y_{n-i} y_{n-j}$*.*

Proof. The independence between $\mathbf{x}$ and $\mathbf{b}$ allows us to write:

$$\begin{aligned} \|\mathbf{h} * \mathbf{x}\|^2 &= \mathscr{E}\left\{\|\mathbf{h} * \mathbf{y}\|^2\right\} - \mathscr{E}\left\{\|\mathbf{h} * \mathbf{b}\|^2\right\} \\ &= \mathscr{E}\left\{\|\mathbf{h} * \mathbf{y}\|^2\right\} - N\sigma^2 \qquad (\|\mathbf{h}\|^2 = 1) \end{aligned}$$

The optimal filter $\mathbf{h}_{\text{opt}}$ that minimizes $\|\mathbf{h} * \mathbf{x}\|^2$ is thus given by:

$$\mathbf{h}_{\text{opt}} = \text{argmin}_{\mathbf{h}}\ \|\mathbf{h} * \mathbf{y}\|^2 \text{ subject to } \|\mathbf{h}\|^2 = 1$$

Introducing the Lagrange multiplier λ, we obtain the following (unconstrained) functional to minimize:

$$\begin{aligned} J(\mathbf{h}) &= \|\mathbf{h} * \mathbf{y}\|^2 - \lambda(\|\mathbf{h}\|^2 - 1) \\ &= \sum_{i,j} h_i \underbrace{\sum_{n} y_{n-i} y_{n-j}}_{a_{i,j}} h_j - \lambda(\|\mathbf{h}\|^2 - 1) \\ &= \mathbf{h}^{\text{T}} \mathbf{A} \mathbf{h} - \lambda(\|\mathbf{h}\|^2 - 1) \end{aligned}$$

and thus:

$$\nabla_{\mathbf{h}} J(\mathbf{h}) = 2\mathbf{A}\mathbf{h} - 2\lambda\mathbf{h} = 0 \Longleftrightarrow \mathbf{A}\mathbf{h} = \lambda\mathbf{h} \tag{3.35}$$

We finally demonstrate Proposition 4, by noting that:

$$\mathbf{h}_{\text{opt}}^{\text{T}} \mathbf{A} \mathbf{h}_{\text{opt}} \overset{(3.35)}{=} \lambda_{\text{opt}} \underbrace{\|\mathbf{h}_{\text{opt}}\|^2}_{=1} = \lambda_{\text{opt}} \text{ is minimum } \Longleftrightarrow \lambda_{\text{opt}} = \lambda_{\min}$$

■

[6] Note that the proposed filter design falls within the broader framework of "eigenfilters" design introduced by Vaidyanathan and Nguyen in [104].

An estimate of the AWGN variance could be simply obtained as

$$\hat{\sigma}^2 = \frac{1}{N}\|\mathbf{h}_{\text{opt}} * \mathbf{y}\|^2$$

However, in order to improve the overall performances, we propose to apply a more robust estimation procedure on the optimally filtered noisy data. The idea is the following: as observed in Figure 3.12, there are still some sparse structures in the residual image that can bias the estimate of σ. We thus propose to compute the noise standard deviation in every blocks of the residual image. Since most of these local estimates will be close to the actual noise standard deviation, the latter can be estimated as the mode of the smoothed histogram of the local noise standard deviations. An overview of the proposed *hybrid* estimation procedure is given in Figure 3.12. In practice, we found that a good trade-off between estimation accuracy and variance can be achieved with a 3×3 optimized filter $\mathbf{h}_{\text{opt}}$ and local estimates of the noise standard deviation computed inside blocks of size 25×25.

As observed in Table 3.3, the proposed estimation procedure is more accurate than the widely-used wavelet-domain MAD estimator, while having a comparable variance. Consequently, the PSNR loss ($0-0.4$ dB) due to the error in the estimation of σ is much lower with the proposed approach (as shown in Figure 3.13, when using our interscale wavelet thresholding (3.18)).

3.7 Experiments

3.7.1 Grayscale Images

In this section, we compare the proposed SURE-LET approach with some of the best state-of-the-art wavelet-based techniques: Sendur's *et al.* bivariate MAP estimator with local signal variance estimation [84], Portilla's *BLS-GSM* [81] and Pižurica's *ProbShrink* [82]. In all comparisons, we use a critically sampled orthonormal wavelet basis with eight vanishing moments (*sym8*) over four (resp. five) decomposition stages for 256×256 (resp. 512×512) images.

PSNR comparisons

We have tested the various denoising methods for a representative set of standard 8-bit[7] grayscale images such as *Lena*, *Barbara*, *Boat*, *Al*, *Goldhill* (size 512×512) and *Peppers*, *House* (size 256×256), corrupted by simulated AWGN at eight different powers $\sigma \in [5, 10, 15, 20, 25, 30, 50, 100]$, which corresponds to PSNR decibel values [34.15, 28.13, 24.61, 22.11, 20.17, 18.59, 14.15, 8.13]. The denoising process has been performed over ten different noise realizations for each noise standard deviation and the resulting PSNRs have been averaged over these ten runs. The parameters of

[7]The PSNR has been thus computed as: $\text{PSNR} = 10\log_{10}\frac{255^2}{\text{MSE}}$.

Table 3.3: Comparison between the proposed AWGN standard deviation estimator and the popular wavelet-domain MAD estimator.

σ		5	10	15	20	25	30	50	100
Method	**Measure**	***Cameraman* 256 × 256**							
MAD	**Relative error**	0.26	0.11	0.07	0.04	0.03	0.02	0.01	**0.00**
	Std of the error	0.06	0.10	0.13	0.17	0.22	0.27	0.45	0.89
Proposed	**Relative error**	**0.05**	**0.02**	**0.02**	**0.02**	**0.01**	**0.01**	**0.00**	0.01
	Std of the error	0.08	0.17	0.24	0.28	0.33	0.37	0.56	1.04
Method	**Measure**	***Bridge* 256 × 256**							
MAD	**Relative error**	0.67	0.25	0.13	0.08	0.05	0.04	0.01	**0.00**
	Std of the error	0.08	0.10	0.14	0.18	0.22	0.25	0.41	0.85
Proposed	**Relative error**	**0.27**	**0.10**	**0.08**	**0.06**	**0.04**	**0.03**	**0.00**	0.01
	Std of the error	0.15	0.17	0.34	0.29	0.29	0.33	0.50	0.96
Method	**Measure**	***Barbara* 512 × 512**							
MAD	**Relative error**	0.28	0.12	0.07	0.04	0.03	0.02	0.01	**0.00**
	Std of the error	0.03	0.04	0.06	0.08	0.10	0.12	0.21	0.41
Proposed	**Relative error**	**0.11**	**0.03**	**0.02**	**0.01**	**0.01**	**0.01**	**0.00**	0.01
	Std of the error	0.08	0.09	0.14	0.17	0.17	0.18	0.20	0.46
Method	**Measure**	***Mandrill* 512 × 512**							
MAD	**Relative error**	0.83	0.32	0.17	0.11	0.07	0.05	0.02	**0.00**
	Std of the error	0.03	0.05	0.08	0.09	0.11	0.14	0.21	0.39
Proposed	**Relative error**	**0.22**	**0.08**	**0.07**	**0.05**	**0.04**	**0.03**	**0.01**	**0.00**
	Std of the error	0.10	0.15	0.25	0.20	0.16	0.16	0.21	0.46

Notes: 1. The relative error has been computed as $\frac{|\hat{\sigma}-\sigma|}{\sigma}$.

2. The estimates have been averaged over 100 noise realizations.

each method have been set according to the values given by their respective authors in the corresponding referred papers. Variations in output PSNRs are thus only due to the denoising techniques themselves. This reliable comparison was only possible thanks to the kindness of the various authors who have provided their respective Matlab codes on their personal websites.

Table 3.4 summarizes the results obtained. To the noteworthy exception of *Barbara*, the results obtained by our interscale estimator (3.18) are already competitive with the best techniques available, when considering non-redundant orthonormal transforms. The integration of the intrascale dependencies brings an additional PSNR improvement of around 0.5 dB, except for the image *Barbara*, where the gain is even more significant (over 1 dB).

When looking closer at the results, we observe that:

- The proposed interscale wavelet estimator gives better results than *BiShrink* which integrates both the inter- and the intrascale dependencies (average gain of 0.6 dB).
- The proposed interscale wavelet estimator gives better results than *ProbShrink* which integrates the intrascale dependencies (average gain of 0.4 dB).
- The proposed interscale wavelet estimator obtains similar or sometimes even

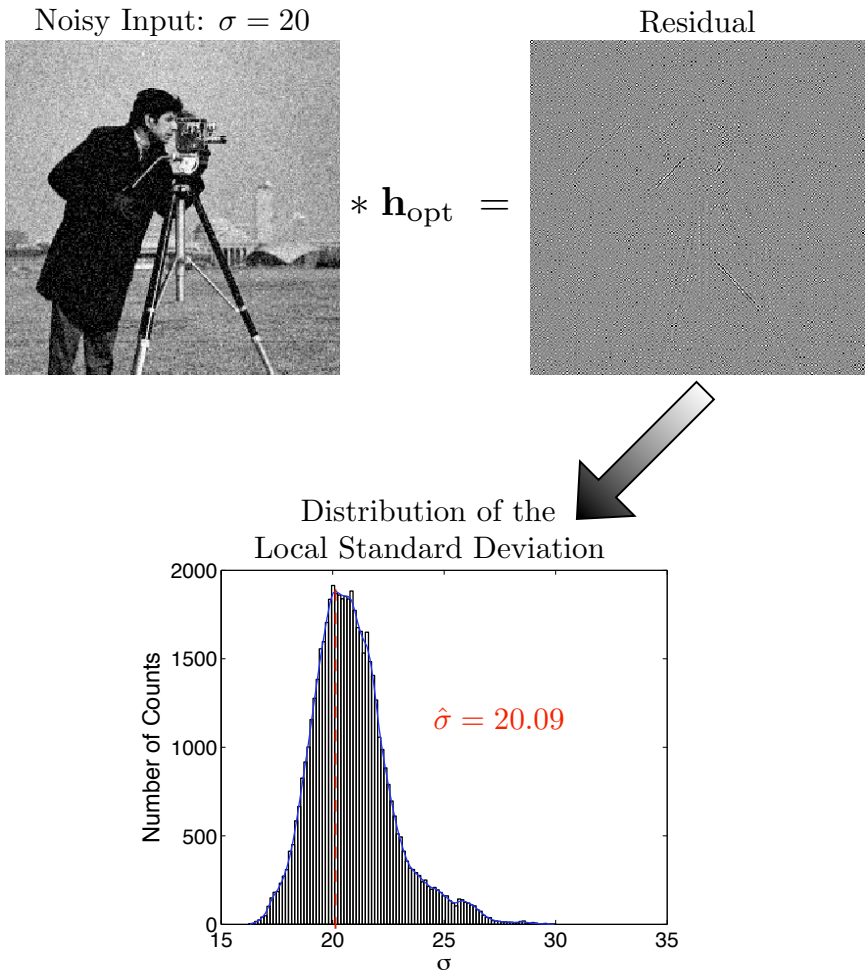


Figure 3.12: Overview of the proposed procedure for estimating the AWGN standard deviation.

better results than *BLS-GSM* for most images.

- The proposed multivariate interscale wavelet thresholding gives consistently higher PSNRs ($0.5 - 1\,\text{dB}$) than the other state-of-the-art wavelet-based denoising algorithms.

Recall that one of the original motivations of this chapter was to design an *orthonormal* wavelet-domain denoising algorithm that would potentially match the quality of the state-of-the-art *redundant* wavelet-domain denoising methods. In Figure 3.14, we thus show a comparison between the proposed multivariate interscale thresholding (3.32) applied in the orthonormal wavelet domain and the *BLS-GSM* applied in two different redundant representations: the classical undecimated wavelet transform and the full steerable (eight orientations per scale) pyramid [47] (FSP). All

Table 3.4: Comparison of some of the most efficient orthonormal wavelet-domain denoising methods.

σ	5	10	15	20	25	30	50	100
Input PSNR	34.15	28.13	24.61	22.11	20.17	18.59	14.15	8.13
Method	***Peppers* 256 × 256**							
BiShrink [84]	36.61	32.56	30.27	28.68	27.48	26.52	23.90	20.78
ProbShrink [82]	36.72	32.69	30.43	28.85	27.65	26.68	23.83	20.85
BLS-GSM [81]	36.80	32.86	30.61	29.07	27.91	26.97	24.39	20.92
Bivariate SURE-LET (3.18)	37.18	33.19	30.91	29.33	28.12	27.14	24.43	21.28
Multivariate SURE-LET (3.32)	**37.40**	**33.51**	**31.31**	**29.78**	**28.61**	**27.66**	**25.00**	**21.50**
Method	***House* 256 × 256**							
BiShrink [84]	37.56	33.64	31.61	30.20	29.11	28.24	25.88	22.89
ProbShrink [82]	37.60	33.84	31.77	30.29	29.20	28.33	26.05	23.17
BLS-GSM [81]	38.04	34.29	32.27	30.83	29.69	28.75	26.20	22.95
Bivariate SURE-LET (3.18)	37.89	34.31	32.34	30.96	29.88	29.00	26.59	23.48
Multivariate SURE-LET (3.32)	**38.45**	**34.90**	**32.93**	**31.53**	**30.43**	**29.53**	**27.02**	**23.66**
Method	**Lena 512 × 512**							
BiShrink [84]	37.57	34.24	32.35	31.01	29.99	29.17	26.91	24.02
ProbShrink [82]	37.64	34.28	32.35	31.01	29.99	29.18	26.97	24.25
BLS-GSM [81]	37.85	34.60	32.69	31.33	30.28	29.43	27.09	24.17
Bivariate SURE-LET (3.18)	37.96	34.56	32.67	31.35	30.34	29.53	27.33	24.61
Multivariate SURE-LET (3.32)	**38.27**	**34.98**	**33.10**	**31.77**	**30.74**	**29.91**	**27.62**	**24.67**
Method	**Barbara 512 × 512**							
BiShrink [84]	36.75	32.52	30.15	28.51	27.29	26.33	23.92	21.53
ProbShrink [82]	36.75	32.48	30.04	28.40	27.20	26.27	23.85	21.60
BLS-GSM [81]	37.05	32.88	30.54	28.92	27.71	26.75	24.26	21.58
Bivariate SURE-LET (3.18)	36.70	32.19	29.66	27.98	26.76	25.83	23.71	21.82
Multivariate SURE-LET (3.32)	**37.46**	**33.35**	**31.01**	**29.38**	**28.15**	**27.18**	**24.70**	**22.07**
Method	**Boat 512 × 512**							
BiShrink [84]	36.19	32.47	30.49	29.10	28.05	27.22	25.07	22.61
ProbShrink [82]	36.21	32.54	30.51	29.11	28.05	27.22	25.13	22.74
BLS-GSM [81]	36.46	32.89	30.89	29.50	28.44	27.59	25.36	22.71
Bivariate SURE-LET (3.18)	36.70	32.90	30.86	29.48	28.45	27.65	25.55	23.09
Multivariate SURE-LET (3.32)	**37.05**	**33.32**	**31.34**	**29.97**	**28.93**	**28.10**	**25.92**	**23.27**
Method	**Al 512 × 512**							
BiShrink [84]	38.00	34.50	32.58	31.25	30.23	29.40	27.12	24.10
ProbShrink [82]	38.11	34.58	32.65	31.24	30.12	29.31	27.18	24.30
BLS-GSM [81]	38.38	34.84	32.95	31.60	30.56	29.71	27.36	24.27
Bivariate SURE-LET (3.18)	38.43	34.90	32.98	31.66	30.66	29.86	27.64	24.67
Multivariate SURE-LET (3.32)	**38.95**	**35.39**	**33.47**	**32.15**	**31.14**	**30.32**	**28.04**	**24.96**
Method	**Goldhill 512 × 512**							
BiShrink [84]	36.18	32.28	30.33	29.07	28.15	27.44	25.59	23.38
ProbShrink [82]	36.07	32.30	30.35	29.07	28.13	27.43	25.63	23.55
BLS-GSM [81]	36.38	32.62	30.69	29.42	28.49	27.75	25.78	23.43
Bivariate SURE-LET (3.18)	36.53	32.69	30.77	29.53	28.62	27.91	26.09	23.94
Multivariate SURE-LET (3.32)	**36.91**	**33.15**	**31.25**	**30.02**	**29.11**	**28.40**	**26.50**	**24.15**

Note: The output PSNRs have been averaged over ten noise realizations.

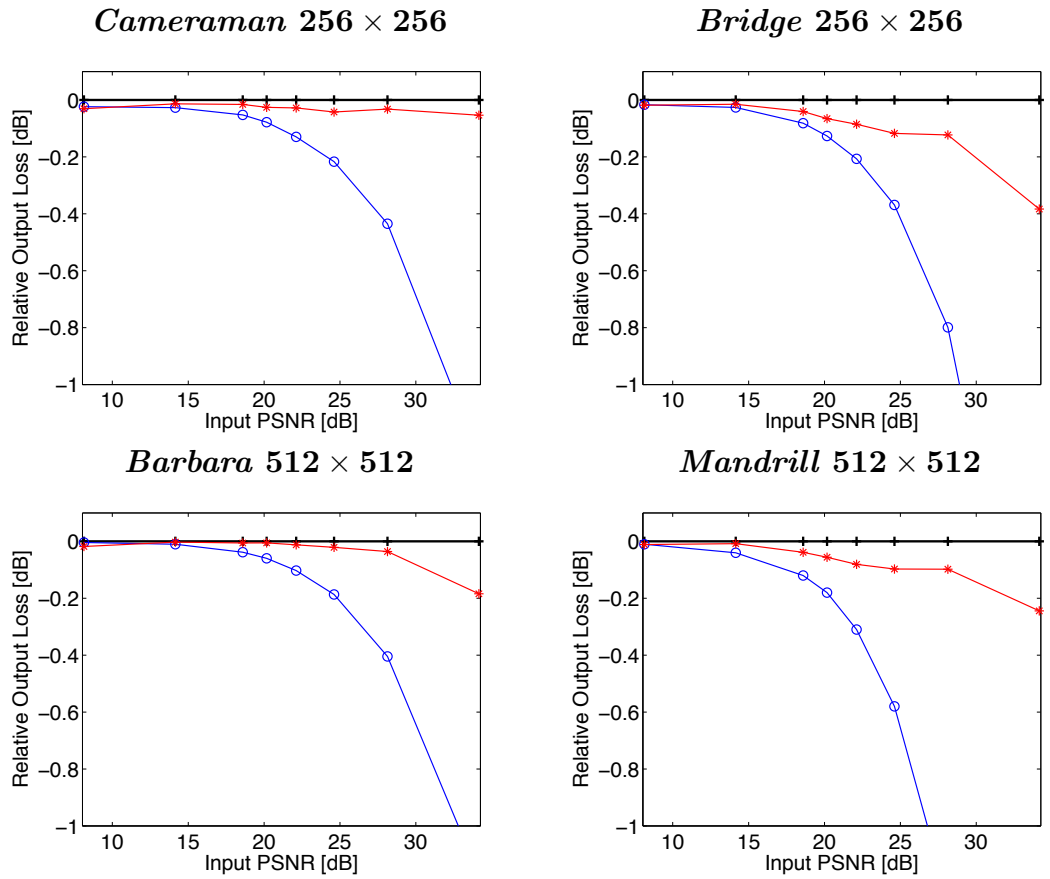


Figure 3.13: PSNR loss due to the error in the estimation of the AWGN standard deviation, when using our interscale thresholding (3.18) with: the actual σ ("+" markers), $\hat{\sigma}$ provided by the wavelet-domain MAD estimator ("$\circ$" markers), and $\hat{\sigma}$ estimated by the proposed procedure ("$*$" markers).

transforms are implemented with periodic boundary extensions. The two wavelet-based transforms use the same *symlet* filters with eight vanishing moments. As observed, we obtain PSNR results similar (± 0.2 dB) to those achieved by the *BLS-GSM* using the UWT. However, the application of the *BLS-GSM* in a FSP outperforms ($+0.5$ dB on average) our non-redundant solution for most images and for most noise levels. A redundant representation allows much more design flexibility than an orthonormal transform. In particular, the better directional selectivity of the FSP leads to a better preservation of both edges and textures. Yet, it is encouraging to notice that we could match the quality of the *BLS-GSM* applied in the standard UWT when considering the same wavelet filters.

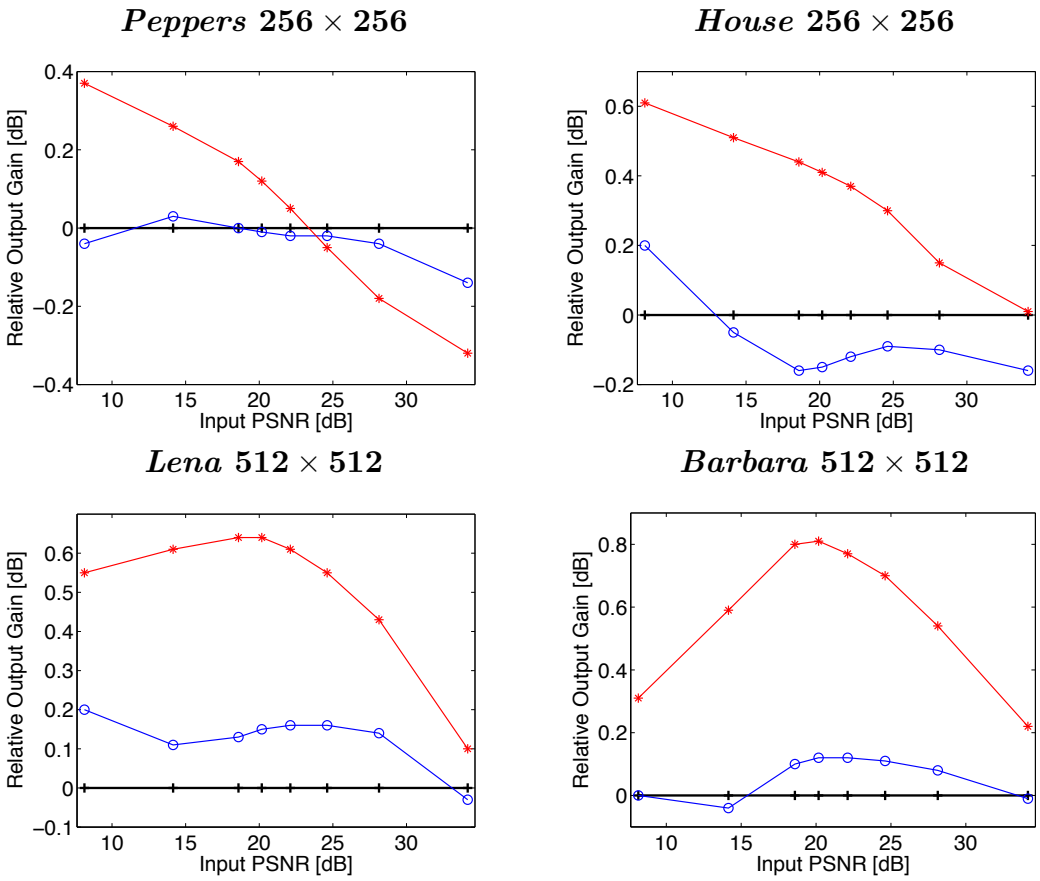


Figure 3.14: Comparison between the proposed multivariate orthonormal wavelet estimator (3.32) ("+" markers) and the *BLS-GSM* applied in two different redundant representations: UWT ("∘" markers) and FSP ("∗" markers).

Visual quality

Although there is no consensual objective way to judge the visual quality of a denoised image, two important criteria are widely used: the visibility of processing artifacts and the conservation of image edges. Processing artifacts usually result from a modification of the spatial correlation between wavelet coefficients (often caused by the zeroing of small neighboring coefficients) and are likely to be reduced by taking into account intrascale dependencies. Instead, edge distortions usually arise from modifications of the interscale coefficient correlations. The amplitude of these modifications is likely to be reduced by a careful consideration of interscale dependencies in the denoising function.

In Figures 3.15, we show the denoised output generated by some of the algorithms

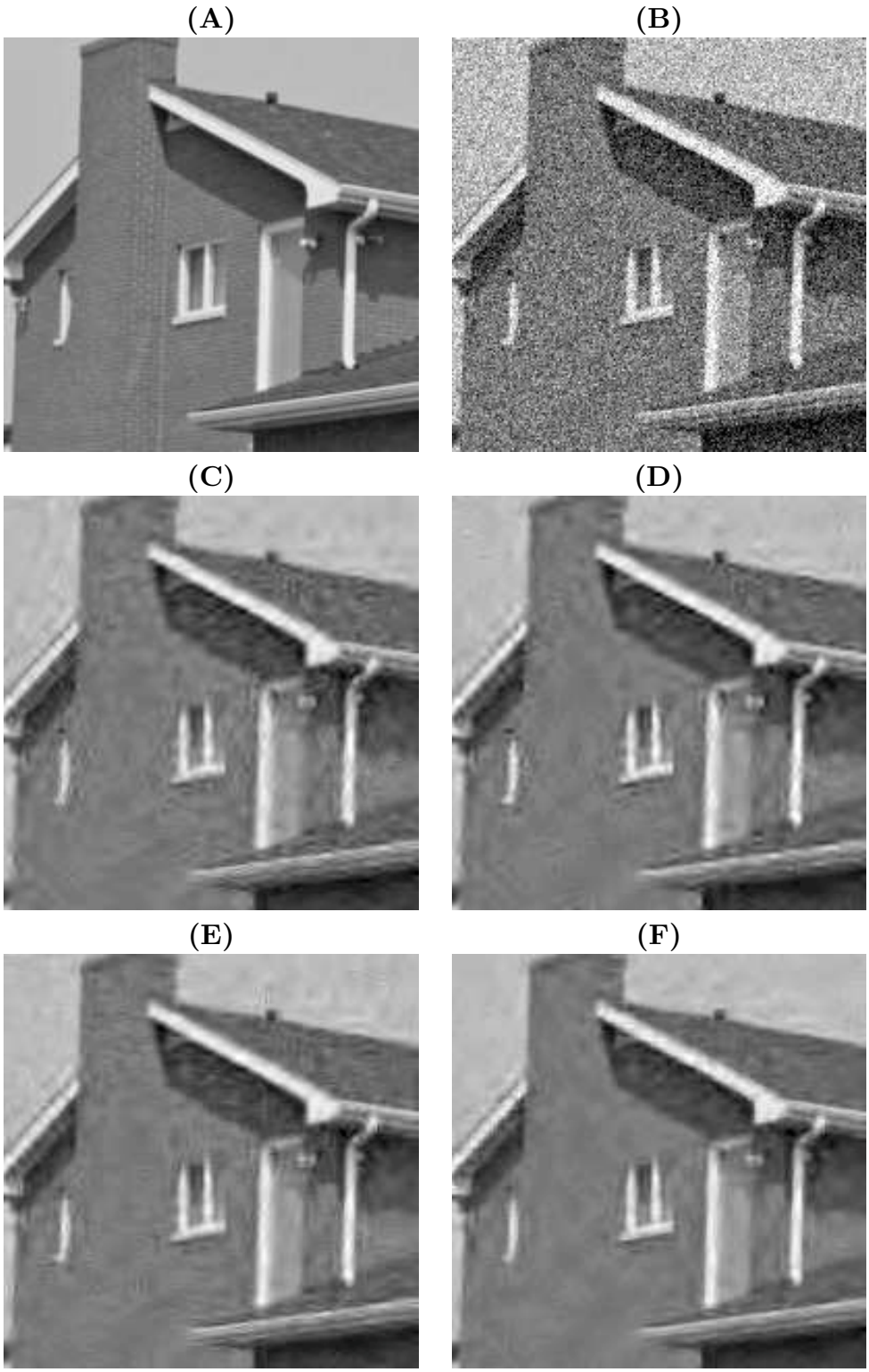


Figure 3.15: **(A)** Part of the noise-free 256×256 *House* image. **(B)** A noisy version of it: SSIM = 0.23. **(C)** Denoised by *BiShrink*: SSIM = 0.73. **(D)** Denoised by *ProbShrink*: SSIM = 0.75. **(E)** Denoised by *BLS-GSM*: SSIM = 0.75. **(F)** Denoised by the proposed multivariate interscale thresholding (3.32): SSIM = 0.78.

compared in this section. We would like to stress that our method exhibits the fewest artifacts, which we attribute to the fact that we are never forcing any wavelet coefficients to zero. We have also reported the SSIM index[8] [5] of the various denoised images: here again, the proposed multivariate interscale thresholding yields the highest SSIM index.

Computation time

It is also interesting to evaluate the various denoising methods from a practical point of view: the computation time. Indeed, the results achieved by overcomplete representations are admittedly superior than the ones obtained by critically sampled wavelet transforms, but their weaknesses are the time they require and their intensive computer memory usage. The execution of the *BLS-GSM* applied in a FSP lasts around 32 s on a MacPro workstation with 2×2.66 GHz Dual-Core for 512×512 images. With our multivariate interscale thresholding (3.18), the whole denoising task takes around 0.18 s for 256×256 images and about 0.68 s for 512×512 images, using the same workstation.

Table 3.5 summarizes the relative computation time of the various methods considered in this section. Note that the main part of the *ProbShrink* and some parts of the *BLS-GSM* are contained in pre-compiled files; this makes their execution a bit faster than the other algorithms which are fully implemented in Matlab.

Table 3.5: Relative computation time of various denoising techniques.

Method	**Normalized Unit of time**	
	256 × 256 images	**512 × 512 images**
Univariate *SURE-LET* (3.11)	1.0	3.9
BiShrink [84]	1.1	4.2
ProbShrink [82]	11.1	45.8
BLS-GSM [81]	15.3	56.7
Bivariate *SURE-LET* (3.18)	2.9	11.6
Multivariate *SURE-LET* (3.32)	6.7	25.1
Redundant *BLS-GSM* [81]	288.1	1193.8

Note: The computation times have been averaged over twenty runs.

[8] For more details on this image quality measure, see Section 1.2.1 of Chapter 1.

3.7.2 Color Images

Color spaces usually consist of $C = 3$ channels and we mostly consider red-green-blue (RGB) representations here. In order to demonstrate the performance of our approach, we assume that the interchannel noise covariance matrix is given by:

$$\mathbf{\Gamma} = \begin{bmatrix} \sigma_R^2 & 0 & 0 \\ 0 & \sigma_G^2 & 0 \\ 0 & 0 & \sigma_B^2 \end{bmatrix}$$

This assumption implies that, in other color spaces, there will usually be noise correlations between the color channels. As an illustration, suppose that we want to perform the denoising in the luminance-chrominance space YUV. An image $\mathbf{y}_{\text{yuv}}$ in YUV is obtained form an original RGB image $\mathbf{y}$ through the following linear transformation:

$$\mathbf{y}_{\text{yuv}} = \underbrace{\begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.1 \end{bmatrix}}_{\mathbf{L}} \mathbf{y} \tag{3.36}$$

The MSE estimate in the YUV color space is finally obtained by replacing $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{\Gamma}$ by respectively $\mathbf{x}_{\text{yuv}} = \mathbf{L}\mathbf{x}$, $\mathbf{y}_{\text{yuv}} = \mathbf{L}\mathbf{y}$ and $\mathbf{\Gamma}_{\text{yuv}} = \mathbf{L}\mathbf{\Gamma}\mathbf{L}^{\text{T}}$ in the expression of the multichannel SURE (3.25).

All the experiments of this section have been carried out on $N = 256 \times 256$ and $N = 512 \times 512$ RGB test images from the set presented in Figure 3.16. We have applied our multichannel interscale thresholding algorithm after 4 or 5 decomposition levels (depending on the size of the image: 256^2 or 512^2) of an orthonormal wavelet transform (OWT) using the standard *sym8* wavelet filter.

Multichannel vs Independent Monochannel Thresholding

Before comparing our results with some of the state-of-the-art denoising procedures, we first want to evaluate the improvements brought by the integration of interchannel dependencies. In Figure 3.17, we compare our multichannel interscale thresholding (3.32) with the interscale thresholding defined in (3.18) applied separately in each channel, both in the standard RGB color space and in the luminance-chrominance space YUV.

As observed, the integration of interchannel dependencies improves the denoising performance considerably, both in the RGB color space (more than +1 dB) and in the more "decorrelated" YUV space (around +0.5 dB). Note that these improvements become even more pronounced (around +1.5–2 dB) when the power of the noise is different in each channel.

Remarkably and contrary to the other denoising algorithms that have been published previously, our results are quite insensitive to the color representation (varia-

Figure 3.16: Test images used in the experiments, referred to as *Image 1* to *Image 8* (numbered from left-to-right and top-to-bottom).

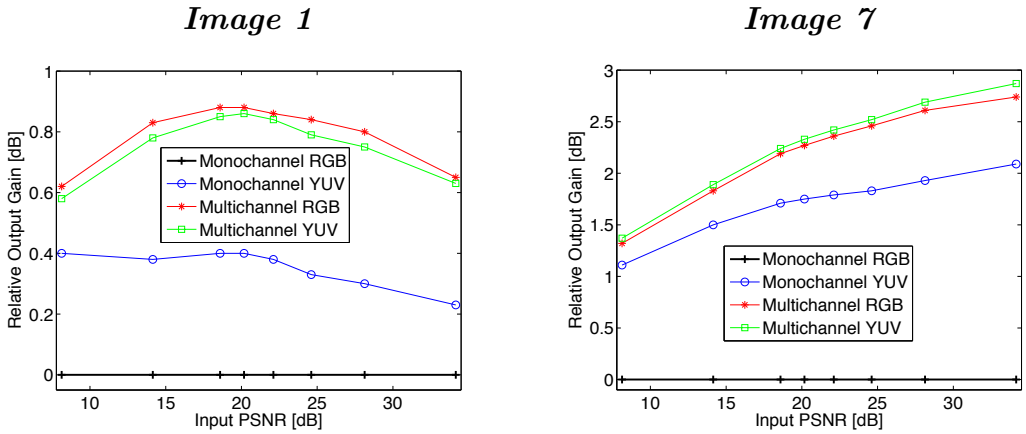


Figure 3.17: PSNR improvements brought by our multichannel strategy, compared to the worst case (monochannel SURE-LET in RGB).

tions of ± 0.1 dB). Indeed, the parameters $\mathbf{a}_k$ of (3.32) can be understood as statistically optimized (in the minimum SURE sense) linear color space transformations in each wavelet zone. From now on, we will thus apply our multichannel algorithm in the RGB color space only.

PSNR Comparisons

We have chosen to compare our method with two state-of-the-art multiresolution-based denoising algorithms:

- Pižurica's *et al.* *ProbShrink-MB* [82], which is a multiband extension of the original grayscale denoiser of the same authors. For color image denoising, it has to be applied in the standard RGB representation, and for equal noise variance in each channel. We have applied this algorithm with a non-redundant orthonormal wavelet transform, as well as with the *undecimated* wavelet tranform (UWT); we have considered the same number of decomposition levels and the same wavelet (*sym8*) as with our method. Since this algorithm has been shown in [82,105] to favorably compare with the multiband wavelet thresholding described in [106], as well as with the vector-based linear MMSE estimator proposed in [107], it constitutes a good reference for evaluating our solution.

- Portilla's *et al.* *BLS-GSM* [81]: although this algorithm has not been designed for multichannel denoising, this is currently the most efficient multiresolution-based grayscale denoiser we know of. For color image denoising, we have simply applied the *BLS-GSM* independently in each RGB channel, considering the *redundant* FSP representation.

  Note that, in all likelihood, a more complete (i.e. including the modeling of local neighborhoods and parents) multichannel extension of the *BLS-GSM* than the one recently initiated by Scheunders *et al.* in [108], would certainly give substantially better results than the independent application of the original *BLS-GSM* that we propose to use here.

In the first experiment, we have corrupted the test images with AWGN having the same variance in each RGB channel. The PSNR results are displayed in Table 3.6. Using the same orthonormal wavelet transform, our multichannel *SURE-LET* algorithm clearly outperforms (often by more than +1 dB) the *ProbShrink-MB*. Despite being performed in a *non-redundant* wavelet representation, our solution gives even better (average gain of nearly +0.5 dB) output PSNRs than the *ProbShrink-MB* applied in the *undecimated* wavelet representation, and similar results to *BLS-GSM* for all the tested images, as well as for the whole range of input noise levels. From a visual point of view, our algorithm holds its own against the best redundant approaches based on multiresolution (see Figure 3.18).

In a second experiment, the test images have been corrupted with AWGN having a different power in each RGB channel. As a comparison, we have used another version of Pižurica's *et al.* *ProbShrink* [105] which is an application of their original grayscale denoiser in the luminance-chrominance color space in the undecimated wavelet transform, consequently referred to as the *UWT ProbShrink-YUV*. The PSNR results are displayed in Table 3.7. We have also reported in this table the results published in [109]. Their algorithm is developed in an orthonormal wavelet

Table 3.6: Comparison of color image denoising algorithms for the same noise level in each RGB channel.

$\sigma_R = \sigma_G = \sigma_B$	5	10	15	20	25	30	50	100
Input PSNR [dB]	34.15	28.13	24.61	22.11	20.17	18.59	14.15	8.13
Method	***Image 1* 256 × 256**							
OWT ProbShrink-MB [82]	36.65	33.18	31.32	29.98	28.94	28.07	25.61	22.69
OWT SURE-LET	**37.91**	**34.46**	**32.60**	31.29	30.25	29.39	26.95	23.73
UWT ProbShrink-MB [82]	37.69	34.22	32.30	30.94	29.96	29.06	26.55	23.72
FSP BLS-GSM [81]	37.57	34.20	32.52	**31.31**	**30.34**	**29.52**	**27.21**	**24.12**
Method	***Image 2* 256 × 256**							
OWT ProbShrink-MB [82]	34.37	30.01	28.17	26.93	26.00	25.25	23.24	20.72
OWT SURE-LET	**35.40**	**31.22**	**29.24**	**27.98**	**27.07**	**26.34**	**24.38**	**21.76**
UWT ProbShrink-MB [82]	35.31	31.21	29.22	27.92	26.99	26.21	24.17	21.65
FSP BLS-GSM [81]	35.35	31.01	29.09	27.91	27.04	26.34	24.37	21.74
Method	***Image 3* 512 × 512**							
OWT ProbShrink-MB [82]	36.37	33.45	31.78	30.59	29.67	28.92	26.88	24.28
OWT SURE-LET	**37.80**	**34.64**	**33.02**	**31.90**	**31.04**	**30.33**	**28.35**	**25.66**
UWT ProbShrink-MB [82]	37.46	34.42	32.69	31.47	30.61	29.83	27.76	25.03
FSP BLS-GSM [81]	37.29	34.45	32.90	31.78	30.89	30.15	28.09	25.40
Method	***Image 4* 512 × 512**							
OWT ProbShrink-MB [82]	35.48	32.49	31.02	29.98	29.14	28.41	26.20	23.55
OWT SURE-LET	**36.62**	**33.35**	31.79	30.72	29.89	29.19	27.16	24.48
UWT ProbShrink-MB [82]	36.33	**33.35**	31.81	30.74	29.96	29.20	26.85	24.28
FSP BLS-GSM [81]	36.34	33.26	**31.89**	**30.92**	**30.13**	**29.46**	**27.47**	**24.73**
Method	***Image 5* 512 × 512**							
OWT ProbShrink-MB [82]	33.86	28.90	26.44	24.87	23.73	22.89	20.94	19.31
OWT SURE-LET	**35.12**	**30.49**	**28.15**	**26.64**	**25.55**	**24.71**	**22.59**	**20.37**
UWT ProbShrink-MB [82]	34.83	30.15	27.72	26.17	25.04	24.16	21.98	19.81
FSP BLS-GSM [81]	35.01	30.13	27.66	26.08	24.95	24.07	21.92	19.89
Method	***Image 6* 512 × 512**							
OWT ProbShrink-MB [82]	37.58	34.03	32.01	30.64	29.29	28.24	25.90	23.30
OWT SURE-LET	**39.11**	**35.70**	**33.71**	**32.29**	**31.19**	**30.29**	**27.77**	**24.77**
UWT ProbShrink-MB [82]	38.78	35.23	33.20	31.80	30.77	29.81	26.87	23.97
FSP BLS-GSM [81]	38.40	35.01	33.09	31.74	30.69	29.84	27.47	24.45
Method	***Image 7* 512 × 512**							
OWT ProbShrink-MB [82]	35.47	31.25	29.15	27.76	26.70	25.87	23.77	21.44
OWT SURE-LET	**38.69**	**34.24**	**31.87**	**30.29**	**29.10**	**28.15**	**25.63**	**22.72**
UWT ProbShrink-MB [82]	37.05	32.64	30.36	28.88	27.80	26.87	24.58	21.91
FSP BLS-GSM [81]	36.36	32.17	30.14	28.86	27.92	27.18	25.13	22.50
Method	***Image 8* 512 × 512**							
OWT ProbShrink-MB [82]	39.47	35.97	33.87	32.05	30.85	29.98	27.76	25.00
OWT SURE-LET	**41.05**	**37.56**	**35.49**	**34.00**	**32.84**	**31.88**	**29.26**	26.11
UWT ProbShrink-MB [82]	40.49	36.92	34.82	33.36	32.25	31.24	28.61	25.55
FSP BLS-GSM [81]	40.16	37.03	35.11	33.71	32.62	31.72	29.24	**26.18**

Notes: The output PSNRs have been averaged over ten noise realizations.

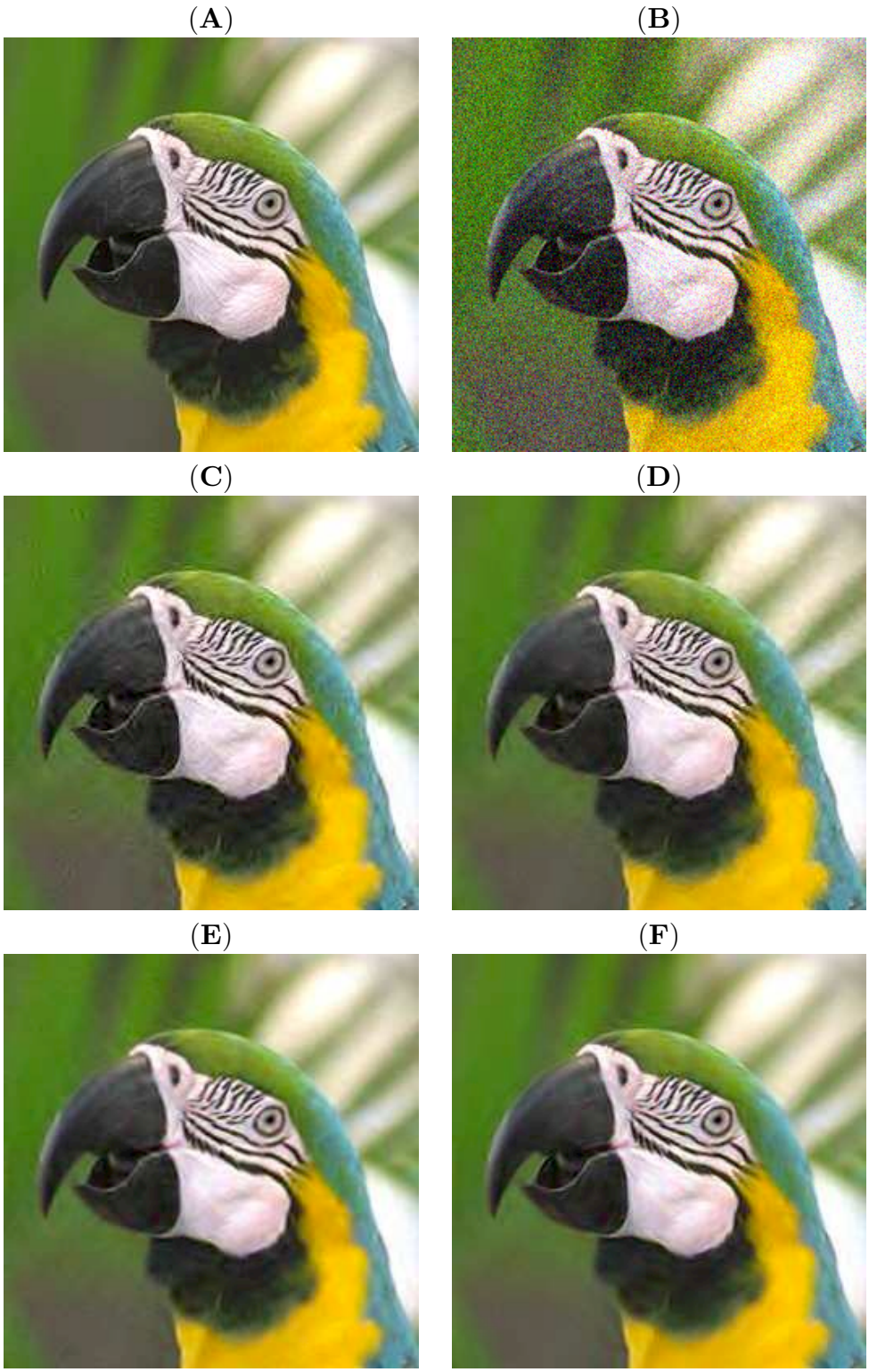


Figure 3.18: (**A**) Part of the noise-free *Image 8*. (**B**) Noisy version of it: PSNR = 20.17 dB. (**C**) Denoised by *ProbShrink-MB* (OWT *sym8*): PSNR = 30.88 dB. (**D**) Denoised by *SURE-LET* (OWT *sym8*): PSNR = 32.83 dB. (**E**) Denoised by *ProbShrink-MB* (UWT *sym8*): PSNR = 32.22 dB. (**F**) Denoised by *BLS-GSM* (FSP): PSNR = 32.60 dB.

transform framework and combines the universal hidden Markov tree (uHMT), a statistical approach devised in [110], with an optimal luminance/color-difference space projection (OCP); it will therefore be referred to as the *OWT uHMT-OCP*. As observed, our multichannel *SURE-LET* approach outperforms these two algorithms in terms of PSNR (almost +1 dB); it even gives better results than the *BLS-GSM* for most images. In Figure 3.19, we show the visual quality of the various algorithms: ours exhibits very few color artifacts, and preserves most of the image details.

Table 3.7: Comparison of color image denoising algorithms for a different noise level in each RGB channel.

$\sigma_R = 38.25$, $\sigma_G = 25.50$, $\sigma_B = 12.75$				
Input PSNR: 19.33 [dB]				
Method	***Image 1***	***Image 2***	***Image 3***	***Image 4***
OWT SURE-LET	**30.63**	**27.19**	**31.41**	29.93
UWT ProbShrink-YUV [105]	29.53	26.55	30.47	29.37
FSP BLS-GSM [81]	30.47	27.18	30.91	**30.13**
Method	***Image 5***	***Image 6***	***Image 7***	***Image 8***
OWT uHMT-OCP [109]	N/A	N/A	29.16	31.46
OWT SURE-LET	**26.12**	**31.34**	**30.09**	**33.17**
UWT ProbShrink-YUV [105]	24.93	29.95	28.25	32.03
FSP BLS-GSM [81]	25.10	30.45	27.83	32.45

Notes: The output PSNRs have been averaged over ten noise realizations.

Finally, we must emphasize that the execution of the un-optimized Matlab implementation of our algorithm only lasts around 1.8s for 512×512 color images on a MacPro workstation with 2×2.66 GHz Dual-Core. To compare with, the best *ProbShrink* requires approximately 6s under the same conditions, whereas the *BLS-GSM* requires about 108s. Besides achieving very competitive denoising results, the proposed solution is also faster than most state-of-the-art algorithms: the interested reader may wish to check these claims with our online demo available at `http://bigwww.epfl.ch/demo/suredenoising-color/index.html`. Not only is it faster, but it is also much more memory effective because it makes use of a non-redundant transformation, an approach that could prove even more valuable for the processing of higher-dimensional data, in particular, tridimensional and moving pictures.

3.7.3 Multichannel Images

Our multichannel SURE-LET algorithm is particularly well-suited to the denoising of multiband images, such as satellite images, and more generally, any stack of images with significant common content (e.g., consecutive moving images or consecutive volume slices). Indeed, thanks to the SURE-based optimization of the linear

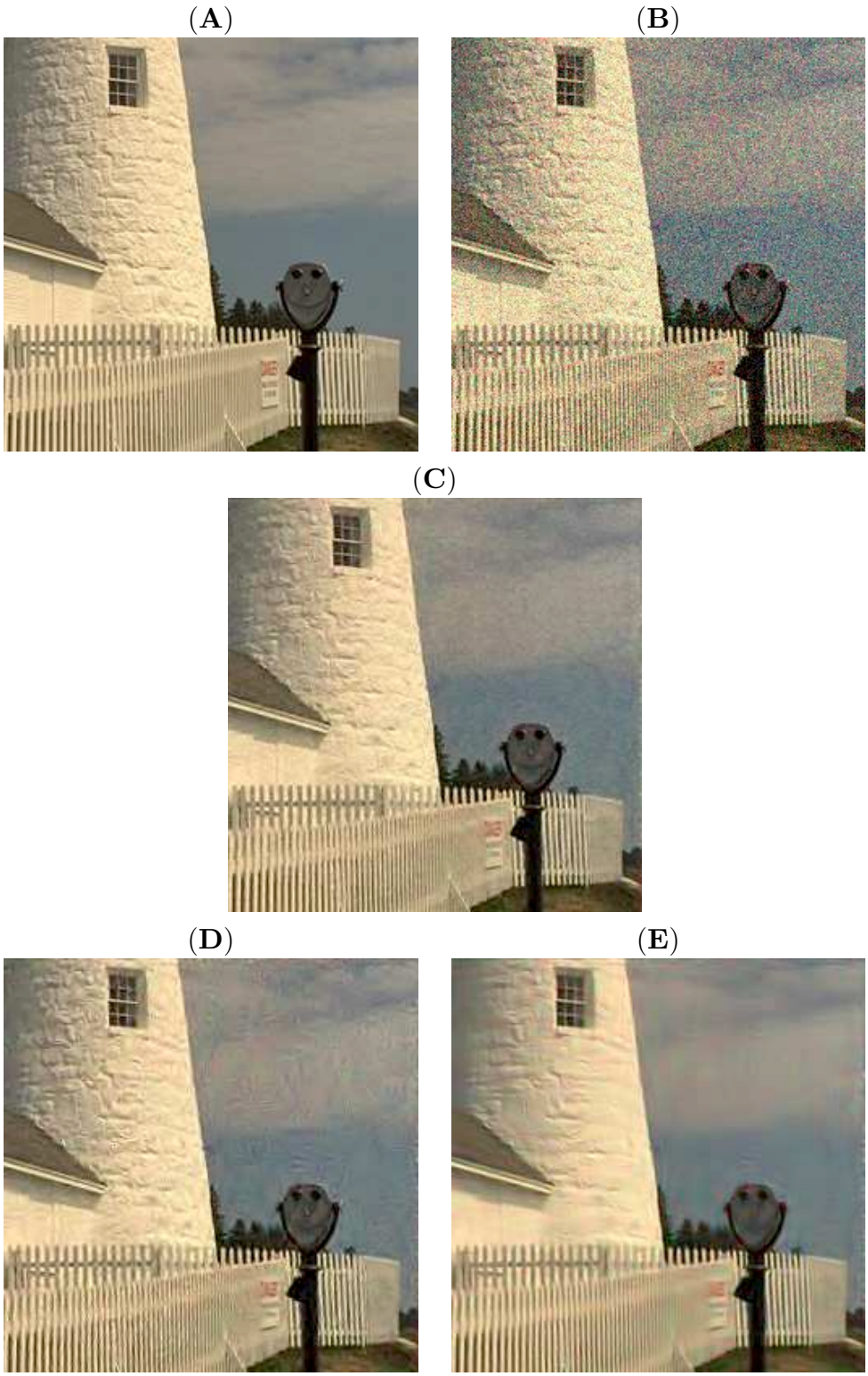


Figure 3.19: (**A**) Part of the noise-free *Image 7*. (**B**) Noisy version of it: PSNR = 19.33 dB ($\sigma_R = 38.25$, $\sigma_G = 25.50$ and $\sigma_B = 12.75$). (**C**) Denoised by our multichannel *SURE-LET* (OWT *sym8*): PSNR = 30.05 dB. (**D**) Denoised by *ProbShrink-YUV* (UWT *sym8*): PSNR = 28.27 dB. (**E**) Denoised by *BLS-GSM* (FSP): PSNR = 27.83 dB.

parameters, the potentially strong similarities between the various channels are efficiently (and automatically) taken into account. There is thus no need to decorrelate the bands beforehand.

For the experiments, we have used two different 7-bands *Landsat* images[9]:

- the first one covers the inland city of Wagga Wagga in Australia. The coverage area shown in Figure 3.20(A) is approximately 15 km by 15 km with a resolution of 30 m (image size of $N = 512 \times 512 \times 7$).
- the second one shows a part of a scene taken over Southern California, encompassing the region from Long Beach to San Diego. The coverage area shown in Figure 3.20(B) is also approximately 15 km by 15 km with a resolution of 30 m (image size of $N = 512 \times 512 \times 7$).

(**A**) (**B**)

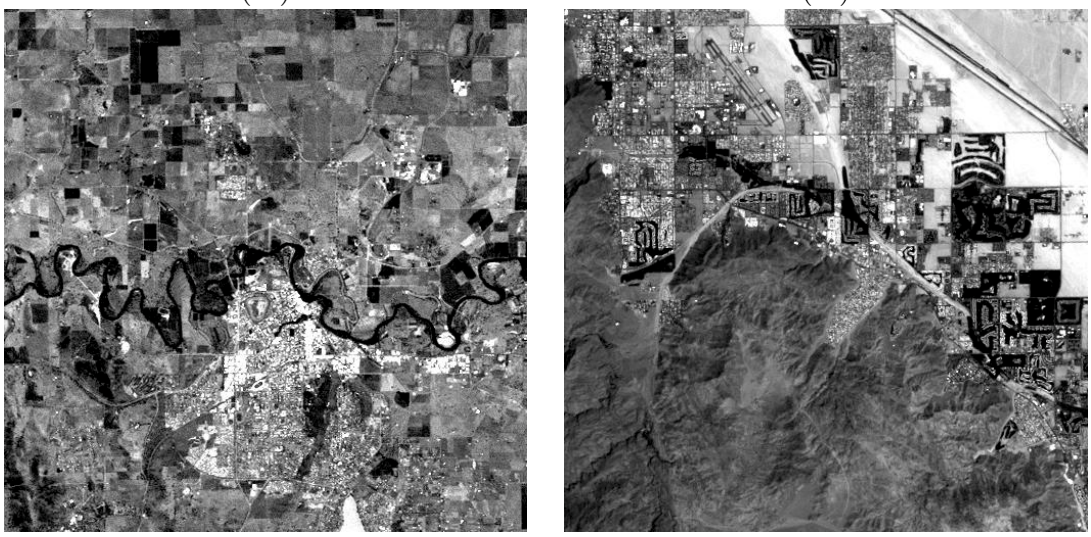


Figure 3.20: (**A**) First band of a *Landsat* image of Wagga Wagga. (**B**) First band of a *Landsat* image showing a part of Southern California.

For the denoising experiments, we have disregarded band 6 of both *Landsat* images, since it is very different from the others (a thermal infrared channel at lower resolution); our test data are therefore of size $N = 512 \times 512 \times 6$. Unfortunately, we were unable to compare our results with the ones obtained by other algorithms specifically devised to handle more than 3 bands because we could neither get the test data used in their experiments nor find the corresponding implementations. However, to have a point of comparison we show the results obtained by the *BLS-GSM* applied separately in each bands.

[9]Data by courtesy of the following website:
`http://ceos.cnes.fr:8100/cdrom-00/ceos1/datasets.htm`

Table 3.8: Comparison of multiband denoising algorithms (same noise level in each channel)

$\sigma_i,\ i \in [1,6]$	5	10	15	20	25	30	50	100
Input PSNR [dB]	34.15	28.13	24.61	22.11	20.17	18.59	14.15	8.13
Method	***Wagga Wagga* 512 × 512**							
FSP BLS-GSM [81]	35.06	30.21	27.70	26.09	24.92	24.02	21.76	19.26
OWT SURE-LET	**35.37**	**30.88**	**28.59**	**27.09**	**26.00**	**25.14**	**22.95**	**20.39**
Method	***Southern California* 512 × 512**							
FSP BLS-GSM [81]	35.52	30.83	28.33	26.67	25.47	24.54	22.22	19.71
OWT SURE-LET	**36.57**	**32.19**	**29.78**	**28.15**	**26.96**	**26.03**	**23.64**	**20.92**

Notes: The output PSNRs have been averaged over ten noise realizations.

As it can be observed in Table 3.8, our multichannel *SURE-LET* clearly outperforms (often by more than +1 dB) the *BLS-GSM*, although it is applied in an orthonormal wavelet representation. A visual comparison is also shown in Figure 3.21 for one particular band. From a computational time point of view, there is an obvious interest in considering a non-redundant transformation: the denoising of the 6 bands of a 512×512 *Landsat* image lasts 6.5s with our algorithm, whereas it takes 210 s with *BLS-GSM*.

3.8 Summary

In this chapter, we have presented several SURE-LET estimators for *orthonormal* wavelet domain denoising. While in most thresholding functions (e.g. soft-threshold), the non-linear transition between high and low magnitude wavelet coefficients is managed by one or more *non-linear* parameters, we have proposed here an alternative pointwise thresholding, ruled by two *linear* parameters. Besides being faster to optimize, our solution achieves better PSNR results than even the optimal sof-thresholding.

However, the restrictive use of an orthonormal transform requires more sophisticated thresholding functions in order to be competitive with the state-of-the-art approaches. We have thus proposed a rigorous procedure (based on group-delay compensation) to take full advantage of the *interscale* relationships that are naturally present in a wavelet representation. Combining these interscale dependencies with the local consistency of the wavelet coefficients (the so-called *intrascale* dependencies) has significantly improved the performance of our simple *pointwise* thresholding, while maintaining a low computational complexity.

In practice, an image can have several channels (e.g. 3 for color images) that can be potentially highly correlated. The standard approach to deal with such images consists in first de-correlating the data (e.g. color space transformation),

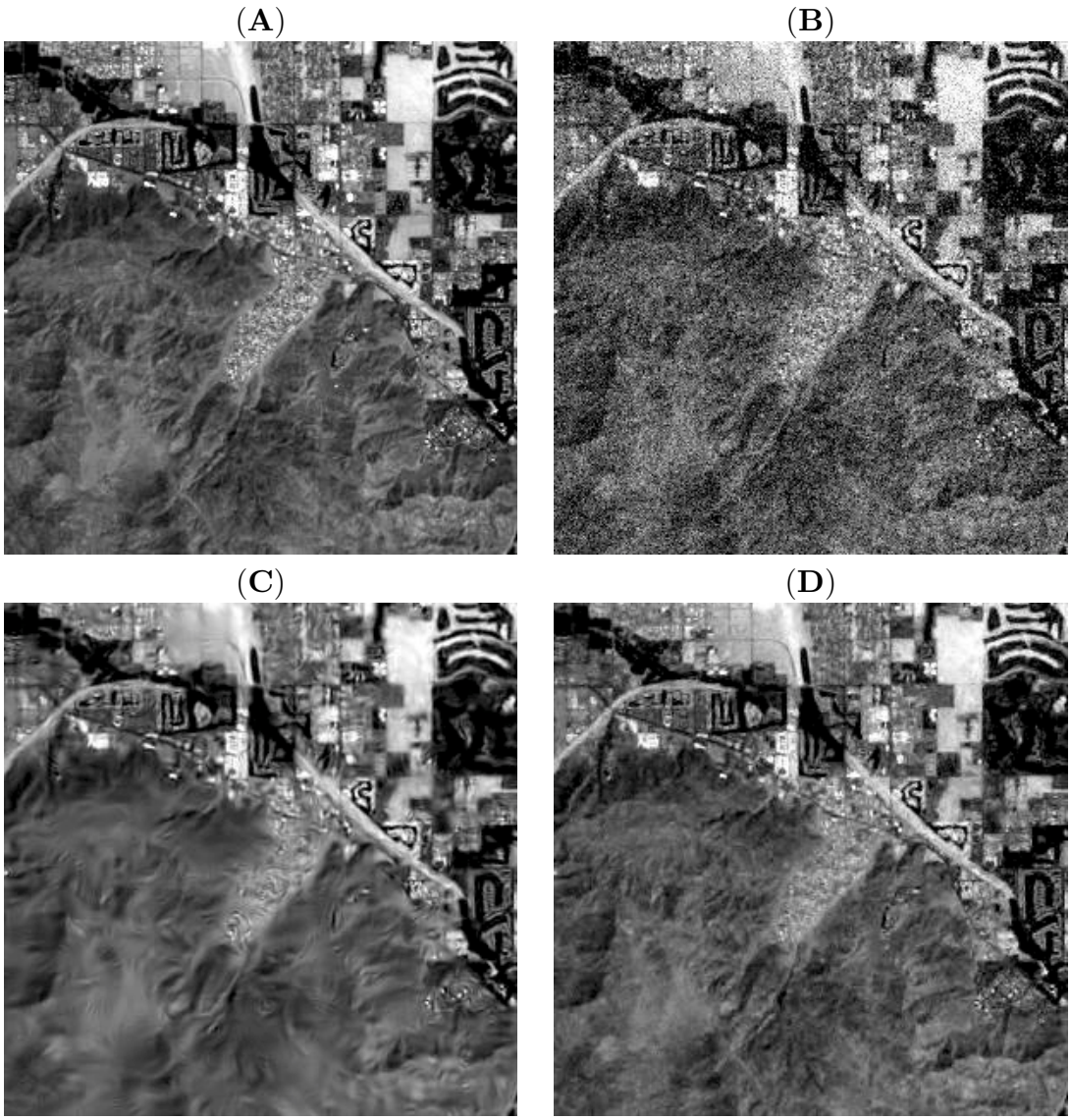


Figure 3.21: (**A**) Part of the first band of the noise-free *Southern California* image. (**B**) Noisy version of it: PSNR = 18.59 dB. (**C**) Denoised by *BLS-GSM* (FSP): PSNR = 24.54 dB. (**D**) Denoised by our multichannel *SURE-LET* (OWT *sym8*): PSNR = 26.03 dB.

and then applying any monochannel denoiser separately on each channel. In this chapter, we have devised a specific *non-separable* wavelet estimator, based on a natural multichannel extension of the orthonormal SURE-LET approach.

Finally, we have assessed the performance of the proposed orthonormal SURE-LET strategy by running several simulations on a representative set of standard

mono- and multichannel images. What comes out of this study is that our solution achieves better performances (both quantitatively and visually) than the state-of-the-art non-redundant denoising methods. In some cases, it can even match the quality of the best multiresolution-based approaches applied in a *redundant* representation, while having a much lighter computational burden.

Chapter 4

SURE-LET in Non-Orthogonal/Redundant Representation

4.1 Motivations

One of the most popular approaches to process noisy images consists in first applying some linear (often redundant) transformation, then performing a nonlinear (and sometimes multivariate) operation on the transformed coefficients, and finally reverting to the image domain by applying the inverse linear transformation. While the choice of the transformation can be justified by invoking well-accepted compression or sparsity arguments (see Section 1.3.4), the nonlinear operation that follows is more frequently based on *ad hoc* statistical hypotheses on the transformed coefficients that are specific to a given algorithm. We have already pointed out that one of the appealing aspects of the SURE-LET toolkit developed so far is that it precisely does *not* require a statistical modelling of the noise-free transformed coefficients.

In this chapter[1], we extend the SURE-LET principle to nonlinear processing performed in an *arbitrary* transformed domain. In particular, we examine the close relations between the choice of transform(s), the design of the transformed-domain nonlinear operation and the global optimization of the whole denoising process. One of the motivations for this study actually originates from the observation made by those who applied soft-thresholding to an undecimated wavelet transform: the *SUREshrink* threshold determination yields substantially worse results than an empirical choice. Unfortunately, this has led some practitioners to wrongly conclude that the SURE approach is unsuitable for redundant transforms, whereas a correct

[1]Some parts of this chapter are based on our published papers [111, 112].

diagnosis should be that it is the independent subband approach that is flawed.

4.2 SURE-LET for Transform-Domain Denoising

Transform-domain denoising consists in first defining a complementary pair of linear transformation operators $\mathscr{D}$ (decomposition) and $\mathscr{R}$ (reconstruction) such that $\mathscr{R}\mathscr{D}$ = Identity: typically, $\mathscr{D}$ is a bank of decimated or undecimated filters. Once the size of the input and output data are frozen, these linear operators are characterized by matrices, respectively $\mathbf{D} = [d_{i,j}]_{(i,j)\in[1;L]\times[1;N]}$ and $\mathbf{R} = [r_{i,j}]_{(i,j)\in[1;N]\times[1;L]}$ that satisfy the *perfect reconstruction* property $\mathbf{RD} = \mathbf{Id}$. In this chapter, we only consider *real* transforms[2], i.e. $d_{i,j}, r_{i,j} \in \mathbb{R}$. The whole denoising process then boils down to the following steps:

1. Apply $\mathbf{D}$ to the noisy signal $\mathbf{y} = \mathbf{x}+\mathbf{b}$ to get the transformed noisy coefficients $\mathbf{w} = \mathbf{D}\mathbf{y} = [w_i]_{i\in[1;L]}$;
2. Apply a (possibly multivariate) thresholding function $\mathbf{\Theta}(\mathbf{w}) = \Big[\theta_i(\mathbf{w})\Big]_{i\in[1;L]}$;
3. Revert to the original domain by applying $\mathbf{R}$ to the thresholded coefficients $\mathbf{\Theta}(\mathbf{w})$, yielding the denoised estimate $\hat{\mathbf{x}} = \mathbf{R}\mathbf{\Theta}(\mathbf{w})$.

Such a denoising procedure can be summarized as a function of the noisy input coefficients:

$$\hat{\mathbf{x}} = \mathbf{F}(\mathbf{y}) = \mathbf{R}\mathbf{\Theta}(\mathbf{D}\mathbf{y}) \tag{4.1}$$

The SURE-LET strategy consists in expressing $\mathbf{F}$ as a *linear expansion* of denoising algorithms $\mathbf{F}_k$, according to:

$$\mathbf{F}(\mathbf{y}) = \sum_{k=1}^{K} a_k \underbrace{\mathbf{R}\mathbf{\Theta}_k(\mathbf{D}\mathbf{y})}_{\mathbf{F}_k(\mathbf{y})}, \tag{4.2}$$

where $\mathbf{\Theta}_k(\cdot)$ are elementary (possibly multivariate) thresholding functions. As already indicated in Section 2.6, retrieving the parameters a_k then boils down to the resolution of the linear system of equations given in Equ. (2.20).

In a general transform-domain SURE-LET framework, the unbiased MSE estimate given in Theorem 1 can be reformulated in the following way:

Corollary 3. *Let* $\mathbf{F}$ *be defined according to* (4.1) *where* $\mathbf{\Theta}$ *denotes a (possibly multivariate) thresholding. Then the MSE between the original and the denoised signal is* unbiasedly *estimated by the following random variable:*

$$\epsilon \;=\; \frac{1}{N}\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + \frac{2\sigma^2}{N}\mathrm{trace}\left\{\mathbf{D}\mathbf{R}\mathbf{J}_{\mathbf{\Theta}}(\mathbf{w})\right\} - \sigma^2 \tag{4.3}$$

[2]We refer the reader to [113] for a recent adaptation of the SURE-LET strategy to complex transforms.

where: $\mathbf{J}_{\mathbf{\Theta}}(\mathbf{w}) = \left[\dfrac{\partial \theta_k(\mathbf{w})}{\partial w_l}\right]_{1\leq k,l\leq L}$ *is the Jacobian matrix of* $\mathbf{\Theta}(\mathbf{w})$.

Proof. By applying Theorem 1, we only have to prove that in the transform-domain SURE-LET framework:

$$\operatorname{div}\{\mathbf{F}(\mathbf{y})\} = \operatorname{trace}\{\mathbf{D}\mathbf{R}\mathbf{J}_{\mathbf{\Theta}}(\mathbf{w})\} \tag{4.4}$$

By using the reconstruction formula $\mathbf{F}(\mathbf{y}) = \mathbf{R}\mathbf{\Theta}(\mathbf{w})$, i.e. $f_n(\mathbf{y}) = \sum_{l=1}^{L} r_{n,l}\theta_l(\mathbf{w})$, and the decomposition formula $\mathbf{w} = \mathbf{D}\mathbf{y}$, i.e. $w_l = \sum_{k=1}^{N} d_{l,k}y_k$, we can successively write the following equalities:

$$\begin{aligned}
\operatorname{div}\{\mathbf{F}(\mathbf{y})\} &= \sum_{n=1}^{N} \frac{\partial f_n(\mathbf{y})}{\partial y_n} \\
&= \sum_{n=1}^{N}\sum_{l=1}^{L} r_{n,l} \sum_{k=1}^{L} \frac{\partial \theta_l(\mathbf{w})}{\partial w_k}\frac{\partial w_k}{\partial y_n} \\
&= \sum_{n=1}^{N}\sum_{l=1}^{L} r_{n,l} \sum_{k=1}^{L} \frac{\partial \theta_l(\mathbf{w})}{\partial w_k} d_{k,n} \\
&= \sum_{k=1}^{L}\sum_{l=1}^{L} \frac{\partial \theta_l(\mathbf{w})}{\partial w_k} \underbrace{\sum_{n=1}^{N} d_{k,n} r_{n,l}}_{[\mathbf{D}\mathbf{R}]_{k,l}} \\
&= \sum_{k=1}^{L}\sum_{l=1}^{L} [\mathbf{D}\mathbf{R}]_{k,l}[\mathbf{J}_{\mathbf{\Theta}}(\mathbf{w})]_{l,k}
\end{aligned} \tag{4.5}$$

and finally conclude that $\operatorname{div}\{\mathbf{F}(\mathbf{y})\} = \operatorname{trace}\{\mathbf{D}\mathbf{R}\mathbf{J}_{\mathbf{\Theta}}(\mathbf{w})\}$. ■

As it appears in this corollary, the computation of the divergence term, i.e. of $\operatorname{trace}\{\mathbf{D}\mathbf{R}\mathbf{J}_{\mathbf{\Theta}}(\mathbf{w})\}$, is a crucial point. In Sections 4.3 and 4.4, we further develop this divergence term for two particular transform-domain processing. We refer the reader to [114] for a general Monte-Carlo technique to evaluate this divergence term for an arbitrary non-linear processing (in particular when $\mathbf{F}$ is not explicitly known).

4.3 Pointwise Estimator

In this section, we only consider a transform-domain *pointwise* processing, i.e. $\mathbf{\Theta}(\mathbf{w}) = \Big[\theta_l(w_l)\Big]_{l\in[1;L]}$. In this particular case, SURE is given by:

Corollary 4. *Let* $\mathbf{F}$ *be defined according to* (4.1) *where* $\boldsymbol{\Theta}$ *denotes a* pointwise *thresholding. Then, the MSE between the original and the denoised signal is* unbiasedly *estimated by the following random variable:*

$$\epsilon = \frac{1}{N}\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + \frac{2\sigma^2}{N}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'(\mathbf{D}\mathbf{y}) - \sigma^2 \tag{4.6}$$

where:

- $\boldsymbol{\alpha} = \operatorname{diag}\{\mathbf{DR}\} = [[\mathbf{DR}]_{l,l}]_{1\leq l\leq L}$ *is a vector made of the diagonal elements of the matrix* $\mathbf{DR}$*;*
- $\boldsymbol{\Theta}'(\mathbf{D}\mathbf{y}) = \boldsymbol{\Theta}'(\mathbf{w}) = [\theta_l'(w_l)]$.

In particular, when $\mathbf{D} = [\mathbf{D}_1^{\mathrm{T}}\ \mathbf{D}_2^{\mathrm{T}}\dots\mathbf{D}_J^{\mathrm{T}}]^{\mathrm{T}}$ *and* $\mathbf{R} = [\mathbf{R}_1\ \mathbf{R}_2\ \dots\ \mathbf{R}_J]$ *where* $\mathbf{D}_i, \mathbf{R}_i$ *are* $N_i \times N$ *and* $N \times N_i$ *matrices, then* $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^{\mathrm{T}}, \boldsymbol{\alpha}_2^{\mathrm{T}}, \dots, \boldsymbol{\alpha}_J^{\mathrm{T}}]^{\mathrm{T}}$ *where* $\boldsymbol{\alpha}_i = \operatorname{diag}\{\mathbf{D}_i\mathbf{R}_i\}$.

The above result is deduced from Corollary 3 by direct calculation.

4.3.1 Case of Non-redundant Transforms

Here, we consider the case of *non-redundant* transforms (i.e. the number of samples is preserved in the transformed domain), and more precisely:

- $\mathbf{D}$ is a full rank matrix of size $N \times N$;
- $\mathbf{R}$ is also a full rank matrix of size $N \times N$.

Then, the divergence term $\boldsymbol{\alpha}$ of Corollary 4 becomes:

$$\boldsymbol{\alpha} = [\underbrace{1\ 1\dots 1}_{N\ \text{times}}]^{\mathrm{T}} \tag{4.7}$$

Indeed, $\mathbf{DR} = \mathbf{RD} = \mathbf{Id}$, so that $\boldsymbol{\alpha} = \operatorname{diag}\{\mathbf{DR}\} = \operatorname{diag}\{\mathbf{Id}\}$.

If in addition the transformation is *orthonormal*, then the reconstruction matrix is simply the transpose of the decomposition matrix, i.e. $\mathbf{R} = \mathbf{D}^{\mathrm{T}}$. Consequently, in Corollary 4 the SURE becomes:

$$\begin{aligned}
\epsilon &= \frac{1}{N}\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + \frac{2\sigma^2}{N}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'(\mathbf{D}\mathbf{y}) - \sigma^2 \\
&= \frac{1}{N}\|\boldsymbol{\Theta}(\mathbf{D}\mathbf{y}) - \mathbf{D}\mathbf{y}\|^2 + \frac{2\sigma^2}{N}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'(\mathbf{D}\mathbf{y}) - \sigma^2 \qquad \text{(orthogonality of } \mathbf{R}\text{)} \\
&\overset{(4.7)}{=} \frac{1}{N}\sum_{i=1}^{N}\Big((\theta_i(w_i) - w_i)^2 + 2\sigma^2\theta_i'(w_i)\Big) - \sigma^2
\end{aligned}$$

where w_i is the i^{th} component of $\mathbf{w} = \mathbf{D}\mathbf{y}$.

For *orthonormal* transforms, SURE is the sum of the specific MSE estimates for each transformed coefficient w_i. The optimization procedure can thus be performed *separately* in the transformed domain, as detailed in Chapter 3. This is specific to orthonormal transforms: non-redundant biorthogonal transforms do not enjoy this property. Yet, the result given in (4.7) still applies and is actually particularly useful for applying our SURE minimization strategy.

4.3.2 Case of Undecimated Filterbank Transforms

We now consider the case of linear redundant transforms characterized by J analysis filters $\tilde{G}_i(z) = \sum_n \tilde{g}_i[n]z^{-n}$ and J synthesis filters $G_i(z) = \sum_n g_i[n]z^{-n}$ as shown in Figure 4.1.

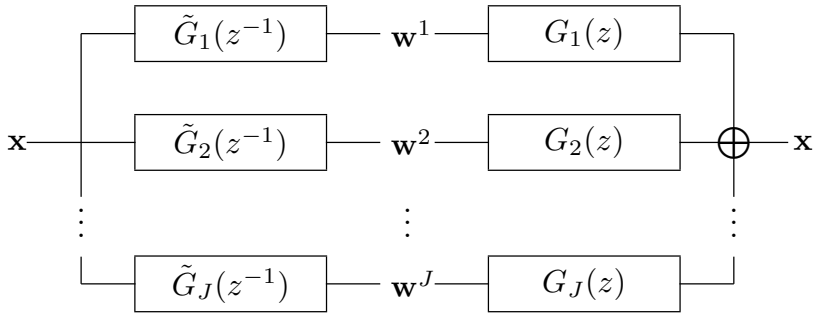


Figure 4.1: Undecimated J-band analysis–synthesis filterbank.

A periodic boundary extension implementation of this structure gives rise to decomposition and reconstruction matrices $\mathbf{D}$ and $\mathbf{R}$ made of J circulant submatrices (i.e. diagonalized with an N-point DFT matrix) $\mathbf{D}_i$ and $\mathbf{R}_i$ of size $N \times N$ each, with coefficients:

$$[\mathbf{D}_i]_{k,l} = \sum_n \tilde{g}_i[l - k + nN]$$
$$[\mathbf{R}_i]_{k,l} = \sum_n g_i[k - l + nN]$$

We then have the following lemma to be used in Corollary 4:

Lemma 2. *When* $\mathbf{D}$ *and* $\mathbf{R}$ *are periodically extended implementations of the analysis–synthesis filterbank of Figure 4.1, the divergence term* $\boldsymbol{\alpha}$ *in* (4.6) *is given by* $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^{\mathrm{T}}\ \boldsymbol{\alpha}_2^{\mathrm{T}}\ \ldots \boldsymbol{\alpha}_J^{\mathrm{T}}]^{\mathrm{T}}$ *where*

$$\boldsymbol{\alpha}_i = \Big(\sum_n \gamma_i[nN]\Big)[\underbrace{1\ 1 \ldots 1}_{N \text{ times}}]^{\mathrm{T}} \tag{4.8}$$

and where $\gamma_i[n]$ *is the* n^{th} *coefficient of the filter* $\tilde{G}_i(z^{-1})G_i(z)$.

Proof. According to Corollary 4 we have to compute $\boldsymbol{\alpha}_i = \text{diag}\,\{\mathbf{D}_i\mathbf{R}_i\}$. Since $\mathbf{D}_i$ and $\mathbf{R}_i$ are circulant matrices the product $\mathbf{D}_i\mathbf{R}_i$ is also circulant and is built using the N-periodized coefficients of the filter $\tilde{G}_i(z^{-1})G_i(z)$, i.e.

$$[\mathbf{D}_i\mathbf{R}_i]_{k,l} = \sum_n \gamma_i[k-l+nN]$$

the diagonal of which yields (4.8). ■

The extension to filterbanks in higher dimensions is straightforward, the summation in (4.8) running over a multidimensional index $\mathbf{n}$.

The filters $\tilde{G}_i$ and G_i usually satisfy the biorthogonality condition

$$\sum_{k=0}^{M_i-1} \underbrace{\tilde{G}_i(z^{-1}e^{-j2\pi k/M_i})G_i(ze^{j2\pi k/M_i})}_{\Gamma_i(ze^{j2\pi k/M_i})} = \lambda_i \tag{4.9}$$

where M_i is a divisor of N. Indeed, undecimated filterbanks are often obtained from critically sampled filterbanks, for which (4.9) holds with $\lambda_i = M_i$. In this case, (4.9) actually specifies the coefficients $\gamma_i[nM_i]$ and thus $\boldsymbol{\alpha}_i = \lambda_i/M_i\,[1\;1\ldots1]^{\mathrm{T}}$.

1. **The Undecimated Wavelet Transform:** A first example of such a transform is the standard undecimated wavelet transform (UWT) which uses $J+1$ ($3J+1$ in two dimensions) orthonormal filters (see Figure 4.2). In that case, the equivalent filters are given by

$$\begin{array}{lllll} \tilde{G}_i(z) & = 2^i G_i(z) & = H(z)H(z^2)\ldots H(z^{2^{i-2}})G(z^{2^{i-1}}) & \text{for } i=1,2,\ldots,J \\ \tilde{G}_{J+1}(z) & = 2^J G_{J+1}(z) & = H(z)H(z^2)\ldots H(z^{2^{J-1}}) & \end{array}$$

   They satisfy (4.9) for $\lambda_i = 1$. This shows that $\boldsymbol{\alpha}_i = 2^{-i}[1\;1\ldots1]^{\mathrm{T}}$ for all $i = 1,2,\ldots J$ and $\boldsymbol{\alpha}_{J+1} = 2^{-J}[1\;1\ldots1]^{\mathrm{T}}$. In a 2D separable framework, these values are extended straightforwardly, thanks to the fact that the 2D filters still satisfy (4.9) for $\lambda_i = 1$: the general result is thus that $\boldsymbol{\alpha}_i$ is given by the (2D) downsampling factor $1/M_i$.

2. **The Overlapping BDCT :** The overlapping M-BDCT is a second example of an undecimated filterbank transform, where the underlying filters are given by $\tilde{G}_i(z) = MG_i(z)$ where:

$$\tilde{g}_i[n] = c_i \cos\left(\frac{\pi(2n+1)i}{2M}\right),\ 0 \le i,n \le M-1 \tag{4.10}$$

   with $c_i = \begin{cases} \dfrac{1}{\sqrt{M}}, \text{ if } i = 0 \\ \dfrac{2}{\sqrt{M}}, \text{ if } i = 1\ldots M-1 \end{cases}$

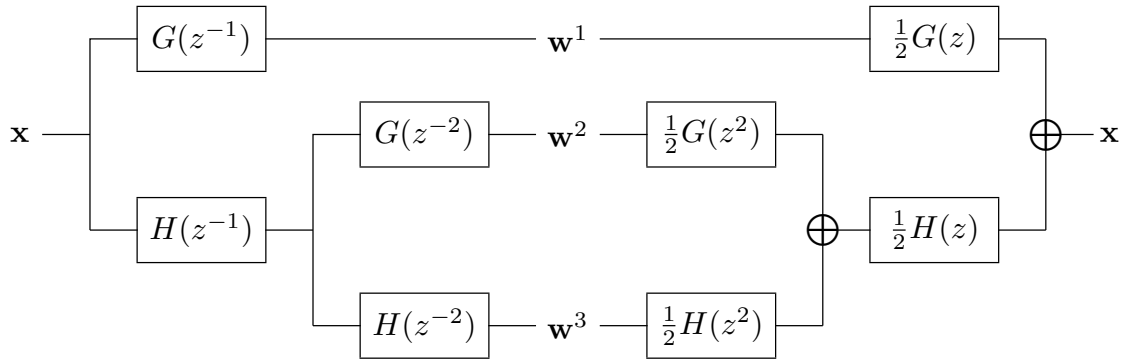


Figure 4.2: The classical undecimated wavelet filterbank for 1D signal. Here, $J = 2$.

The DCT filters also satisfy the biorthogonality condition (4.9) for $\lambda_i = 1$, and consequently $\boldsymbol{\alpha}_i = 1/M\ [1\ 1 \dots 1]^{\mathrm{T}}$ for all $i = 0, 1, \dots M-1$. For d-dimensional signal, the DCT filters are applied separately to each dimension, and thus $\boldsymbol{\alpha}_i = 1/M^d\ [1\ 1 \dots\ 1]^{\mathrm{T}}$.

4.3.3 Pointwise LET

In Section 3.2.2 of Chapter 3, we gave some minimal properties that a pointwise thresholding function should satisfy to be efficient. We propose now a similar subband-dependent denoising function that involves two linear parameters only

$$\theta_j(w; \mathbf{a}_j) = a_{j,1} \underbrace{w}_{t_1(w)} + a_{j,2} \underbrace{w\left(1 - \exp\left(-\left(\frac{w}{3\sigma}\right)^8\right)\right)}_{t_2(w)}, \text{ for } j = 1 \dots J \tag{4.11}$$

which can be seen as a smooth approximation of a *Hard-threshold* (see Figure 4.3). We have experimented with this particular choice of thresholding function, and have found it to be slightly more efficient in a redundant representation than the sum of DOG (3.11) introduced in Section 3.2.2. As in the OWT case, adding more terms in the above LET only brings a marginal denoising gain of $0.1 - 0.2$ dB.

The global image-domain LET can be finally rewritten as

$$\mathbf{F}(\mathbf{y}) \quad = \quad \sum_{j=1}^{J} \sum_{k=1}^{2} a_{j,k} \mathbf{F}_{j,k}(\mathbf{y}) + \underbrace{\mathbf{R}_{J+1} \mathbf{D}_{J+1} \mathbf{y}}_{\text{lowpass}} \tag{4.12}$$

where $\mathbf{F}_{j,k}$ is the image obtained by zeroing all the bands $j' \neq j$ and processing the subband j with the thresholding function $t_k(w)$. Note that, as usual in denoising algorithms, the lowpass residual subband is not processed.

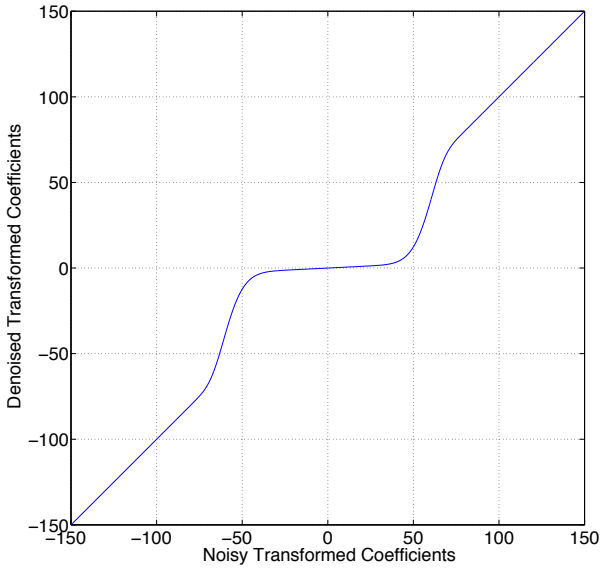


Figure 4.3: A realization of the proposed subband-dependent pointwise thresholding for redundant representation.

The parameters $a_{j,k}$ that minimize the MSE estimate ϵ given in Theorem 1 are then the solution of the linear system of $2J$ equations $\mathbf{Ma} = \mathbf{c}$, where for $\mathbf{k}, \mathbf{l} \in [1; J] \times [1; 2]$,

$$\left\{ \begin{array}{rcl} \mathbf{M} & = & \left[\mathbf{F_k}(\mathbf{y})^{\mathrm{T}} \mathbf{F_l}(\mathbf{y})\right]_{2(k_1-1)+k_2, 2(l_1-1)+l_2} \\ \mathbf{c} & = & \left[\mathbf{F_k}(\mathbf{y})^{\mathrm{T}} (\mathbf{Id} - \mathbf{R}_{J+1}\mathbf{D}_{J+1})\mathbf{y} - \sigma^2 \mathrm{div}\left\{\mathbf{F_k}(\mathbf{y})\right\}\right]_{2(k_1-1)+k_2} \end{array} \right. \tag{4.13}$$

4.3.4 Numerical Computation of the Divergence Term $\boldsymbol{\alpha}$

In the general case where $\mathbf{D}$ and $\mathbf{R}$ are known only by their action on vectors, and not explicitly by their matrix coefficients (typically when only $\mathscr{D}$ and $\mathscr{R}$ are specified), the analytical expression for $\boldsymbol{\alpha}$ may be inconvenient to obtain: in order to build $\boldsymbol{\alpha}$, for each $l = 1, 2, \ldots L$ it is necessary to compute the reconstruction $\mathbf{f}_l = \mathbf{R}\mathbf{e}_l$ (where $[\mathbf{e}_l]_n = \delta_{n-l}$ is the canonical basis of $\mathbb{R}^L$), then the decomposition $\mathbf{D}\mathbf{f}_l$ and keep the l^{th} component. Given that L is of the order of 256^2 (and even much more in the case of redundant transforms) this process may be extremely costly, even considering that it has to be done only once. Fortunately, we can always compute a very good approximation of it using the following numerical algorithm:
For $i = 1 \cdots I$

1. Generate a normalized white Gaussian noise $\mathbf{b}_i \in \mathbb{R}^L$;

2. Apply the reconstruction matrix to $\mathbf{b}_i$ to get the vector $\mathbf{r}_i = \mathbf{R}\mathbf{b}_i$ of size $N \times 1$;

3. Apply the decomposition matrix to $\mathbf{r}_i$ to get the vector $\mathbf{b}'_i = \mathbf{D}\mathbf{R}\mathbf{b}_i$ of size $L \times 1$;

4. Compute the element-by-element product of $\mathbf{b}'_i$ with $\mathbf{b}_i$ to get a vector of L coefficients $\mathbf{v}_i = \operatorname{diag}\left\{\mathbf{b}'_i\mathbf{b}_i^{\mathrm{T}}\right\}$, which can be viewed as a realization of the random vector $\mathbf{v} = \operatorname{diag}\left\{\mathbf{D}\mathbf{R}\mathbf{b}\mathbf{b}^{\mathrm{T}}\right\}$.

end

An approximate value $\hat{\boldsymbol{\alpha}}$ for $\operatorname{diag}\{\mathbf{DR}\}$ is finally obtained by averaging the realizations $\mathbf{v}_i$ over I runs (typically $I = 1000$ provides great accuracy):

$$\hat{\boldsymbol{\alpha}} = \frac{1}{I}\sum_{i=1}^{I}\mathbf{v}_i \tag{4.14}$$

The above algorithm is justified by the following lemma[3]:

Lemma 3. *Let* $\mathbf{b}$ *be a normalized white Gaussian noise with* L *components, and define* $\mathbf{b}' = \mathbf{D}\mathbf{R}\mathbf{b}$*. Then, we have the following equality:*

$$\mathscr{E}\left\{\operatorname{diag}\left\{\mathbf{b}'\mathbf{b}^{\mathrm{T}}\right\}\right\} = \operatorname{diag}\{\mathbf{DR}\} \tag{4.15}$$

Proof.

$$\mathscr{E}\left\{\operatorname{diag}\left\{\mathbf{D}\mathbf{R}\mathbf{b}\mathbf{b}^{\mathrm{T}}\right\}\right\} = \operatorname{diag}\{\mathbf{DR}\underbrace{\mathscr{E}\left\{\mathbf{b}\mathbf{b}^{\mathrm{T}}\right\}}_{\mathbf{Id}}\} = \operatorname{diag}\{\mathbf{DR}\}$$

■

The numerical computation of $\operatorname{diag}\{\mathbf{DR}\}$ can be performed offline for various image sizes, since it does not depend specifically on the image (but for its size) nor on the noise level.

Influence of the Boundary Extensions

One of the main drawbacks of any transform-domain denoising algorithm is the potential generation of boundary artifacts by the transform itself. Decreasing these effects is routinely done by performing boundary extensions, the most popular choice being symmetric extension and periodic extension. Thus, the effect of these extensions boils down to replacing the transformation $\mathbf{D}$ by another transformation, $\mathbf{D}'$.

Indeed, usual boundary extensions are linear pre-processing applied to the available data $\mathbf{y}$ and can therefore be expressed in a matrix form. In particular, for a

[3]Note that a similar result for the stochastic estimation of the trace or the diagonal of a matrix can be found in [115, 116].

given boundary extension of length E, i.e. characterized by an $E \times N$ matrix $\mathbf{H}$, the denoising process becomes:

$$\begin{aligned}\mathbf{F}(\mathbf{y}) &= [\mathbf{Id}_N \ \mathbf{0}_{N\times E}]\, \mathbf{R}_{N+E}\, \mathbf{\Theta} \left(\mathbf{D}_{N+E} \begin{bmatrix} \mathbf{y} \\ \mathbf{H}\mathbf{y} \end{bmatrix} \right) \\ &= \mathbf{R}'\mathbf{\Theta}(\mathbf{D}'\mathbf{y})\end{aligned}$$

where $\mathbf{D}_{N+E}$ (resp. $\mathbf{R}_{N+E}$) is the matrix corresponding to the linear transformation $\mathscr{D}$ (resp. $\mathscr{R}$) when the input signal is of size $N + E$. Any boundary handling can therefore be seen as a modification of the decomposition matrix $\mathbf{D}$ that must be taken into account when computing the divergence term, namely $\operatorname{diag}\{\mathbf{D}'\mathbf{R}'\}$. This is where Lemma 3 is particularly useful: although the implementation of the transformations $\mathscr{D}$ and $\mathscr{R}$ with the adequate boundary extensions may be straightforward, the explicit computation of the coefficients of the matrices $\mathbf{R}'$ and $\mathbf{D}'$ is tedious, and Lemma 3 avoids this computation.

4.3.5 Summary of the algorithm

The proposed SURE-LET algorithm for a pointwise processing in an arbitrary transformed domain can be summarized as follows:

1. Perform a boundary extension on the noisy image;
2. Perform a linear transform on the extended noisy image;
3. **For** $j = 1 \dots J$ (number of bandpass subbands),
   For $k = 1, 2$ (number of thresholds in the LET),
   (a) Apply the *pointwise* thresholding functions t_k defined in (4.11) to the current subband $\mathbf{w}^j$;
   (b) Reconstruct the processed subband by setting all the other subbands to zero to obtain $\mathbf{F}_{j,k}(\mathbf{y})$;
   (c) Compute the first derivative of t_k for each coefficient of the current subband $\mathbf{w}^j$ and build the corresponding coordinate of $\mathbf{c}$ as exemplified by (4.13);

   end
   end
4. Compute the matrix $\mathbf{M}$ given in Equ. (4.13) and deduce the optimal (in the minimum SURE sense) linear parameters $a_{j,k}$'s;
5. The noise-free image $\hat{\mathbf{x}}$ is finally estimated by the sum of each $\mathbf{F}_{j,k}$ weighted by its corresponding SURE-optimized $a_{j,k}$ and the reconstructed lowpass residual subband, as described in Equ. (4.12).

In the particular case of the BDCT, where the number of subbands can be huge (e.g. $(J+1) = 256$ for 16×16 blocks), we propose to process together subbands that are in the same radial frequency band (see Figure 4.4). In practice, we found that considering only $J = 8$ radial frequency bands leads to near optimal PSNR performances, while considerably reducing the overall number of parameters (e.g. from 2×255 to 2×8, no parameters being assigned to the lowpass residual subband).

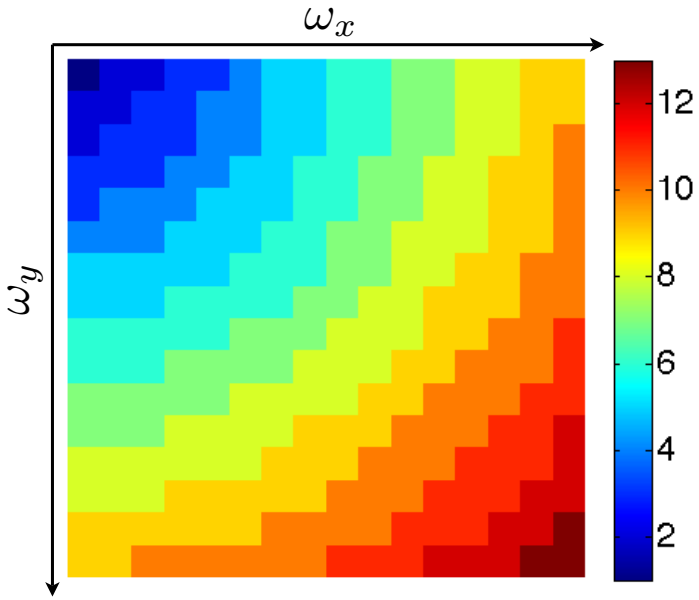


Figure 4.4: BDCT subbands grouping based on their radial frequency: subbands belonging to the same radial frequency band have the same color and share the same LET parameters.

4.3.6 Computational Complexity

The computational complexity of the proposed pointwise SURE-LET algorithm can be evaluated as follows (with N being the number of samples, J the number of processed bandpass subbands and K the overall number of terms in the LET):

1. FFT-based implementation of an undecimated $(J+1)$-band analysis filterbank: $\mathcal{O}((J+1) \cdot N \cdot \log_2(N))$;
2. Application of the pointwise thresholding and of its derivative: $\mathcal{O}(2J \cdot N)$;
3. Reconstruction of each processed subbands and of the lowpass residual: $\mathcal{O}((2J+1) \cdot N \cdot \log_2(N))$;

4. Construction of the *symmetric* $K \times K$ matrix $\mathbf{M}$: $\mathcal{O}(K(K+1)/2 \cdot N)$;

5. Construction of the $K \times 1$ vector $\mathbf{c}$: $\mathcal{O}(K \cdot N)$;

6. Resolution of the linear system of K equations: $\mathcal{O}(K^3)$;

7. Obtention of the final estimate $\hat{\mathbf{x}}$: $\mathcal{O}((K+1) \cdot N)$.

Considering a $N = 256 \times 256$ image and 4 levels of decomposition of the UWT (i.e. $J = 12$ bandpass subbands and thus $K = 24$) leads to an overall number of around $1'000$ operations per pixel. Using an overcomplete 12×12 BDCT (i.e. $J = 143$) and considering only 8 distinct radial frequency bands (i.e. $K = 16$) increases the computational cost to $\sim 7'350$ operations per pixel.

4.4 Multivariate Estimator

In this section, we consider a multivariate processing which takes into account $(2M+1)$ (resp. $(2M+1)^2$ for 2D signal) neighboring transformed coefficients, i.e.

$$\mathbf{\Theta}(\mathbf{w}) = \Big[\theta_l(\mathbf{w}_l)\Big]_{l\in[1;L]}, \text{ where } \mathbf{w}_l = [w_{l-M}\ w_{l-M+1} \dots w_{l+M}]^{\mathrm{T}} \tag{4.16}$$

4.4.1 Multivariate SURE

In the case of a *multivariate* processing applied in an arbitrary transformed-domain, SURE is given by:

Corollary 5. *Let* $\mathbf{F}$ *be defined according to* (4.1) *where* $\mathbf{\Theta}$ *denotes a* multivariate *thresholding as defined in* (4.16). *Then the MSE between the original and the denoised signal is* unbiasedly *estimated by the following random variable:*

$$\epsilon = \frac{1}{N}\|\mathbf{F}(\mathbf{y}) - \mathbf{y}\|^2 + \frac{2\sigma^2}{N}\sum_{m=-M}^{M} \boldsymbol{\alpha}_m^{\mathrm{T}} \mathbf{\Theta}'_m(\mathbf{w}) - \sigma^2 \tag{4.17}$$

where:

- $\boldsymbol{\alpha}_m = \operatorname{diag}\{\mathbf{S}_m\mathbf{DR}\} = [[\mathbf{S}_m\mathbf{DR}]_{l,l}]_{1\le l\le L}$ *is a vector made of the diagonal elements of the matrix* $\mathbf{S}_m\mathbf{DR}$. $\mathbf{S}_m$ *is a* $L \times L$ *permutation matrix which operates a cyclic shift of length* m.

- $\mathbf{\Theta}'_m(\mathbf{w}) = \left[\dfrac{\partial\theta_l(\mathbf{w}_l)}{\partial w_{l-m}}\right]_{1\le l\le L}$.

Proof. Similarly to the proof of Corollary 3, the successive use of the reconstruction and decomposition formulas allows us to write:

$$\begin{aligned}
\operatorname{div}\{\mathbf{F}(\mathbf{y})\} &= \sum_{n=1}^{N} \frac{\partial f_n(\mathbf{y})}{\partial y_n} \\
&= \sum_{n=1}^{N}\sum_{l=1}^{L} r_{n,l} \sum_{m=-M}^{M} \frac{\partial \theta_l(\mathbf{w}_l)}{\partial w_{l-m}} \frac{\partial w_{l-m}}{\partial y_n} \\
&= \sum_{m=-M}^{M}\sum_{l=1}^{L} \frac{\partial \theta_l(\mathbf{w}_l)}{\partial w_{l-m}} \underbrace{\sum_{n=1}^{N} d_{l-m,n} r_{n,l}}_{[\mathbf{S}_m\mathbf{DR}]_{l,l}} \\
&= \sum_{m=-M}^{M} \operatorname{diag}\{\mathbf{S}_m\mathbf{DR}\}^{\mathrm{T}} \boldsymbol{\Theta}'_m(\mathbf{w})
\end{aligned}$$

■

To practically evaluate the terms $\operatorname{diag}\{\mathbf{S}_m\mathbf{DR}\}$, we can simply apply the numerical procedure described in Section 4.3.4 for each shift by m: after Step 3, the vector $\mathbf{b}'_i$ must be further multiplied by the matrix $\mathbf{S}_m$; i.e., its components must be circularly shifted by m positions.

4.4.2 Multivariate LET

In this section, we propose a multivariate thresholding based on the pointwise thresholding $\theta_j(w)$ defined in (4.11). The idea is to use the local energy of the transformed coefficients to coarsely distinguish the potentially high-magnitude coefficients from the potentially low-magnitude ones. Each of the two resulting classes of transformed coefficients is then processed by its corresponding pointwise thresholding $\theta_j(w;\mathbf{a}_j)$ (respectively $\theta_j(w;\mathbf{b}_j)$), leading to the following multivariate thresholding:

$$\theta_j(\mathbf{w};\mathbf{a}_j,\mathbf{b}_j) = \gamma(\|\mathbf{w}\|)\ \theta_j(w;\mathbf{a}_j) + (1-\gamma(\|\mathbf{w}\|))\ \theta_j(w;\mathbf{b}_j) \tag{4.18}$$

Here, $\gamma(x) = e^{-\left(\frac{x}{T}\right)^8}$ is a discriminative function based on the local energy computed inside a $(2M+1)$ neighborhood of transformed coefficients. In our experiments, we found that $T = 2(2M+1)\sigma$ gave the best results.

The general behavior of the proposed multivariate thresholding is the following (see Figure 4.5):

- For low-magnitude local energies (i.e. $\|\mathbf{w}\| \ll T$), $\gamma(\|\mathbf{w}\|)$ is close to 1 and thus $\theta_j(\mathbf{w}) \simeq \theta_j(w;\mathbf{a}_j)$;
- For high-magnitude local energies (i.e. $\|\mathbf{w}\| \gg T$), $\gamma(\|\mathbf{w}\|)$ is close to 0 and thus $\theta_j(\mathbf{w}) \simeq \theta_j(w;\mathbf{b}_j)$.

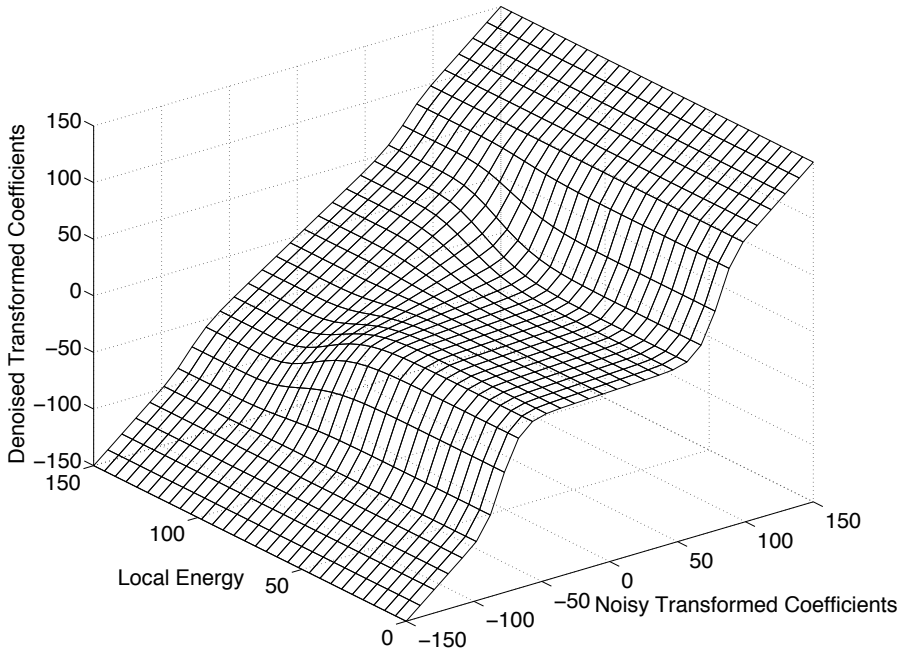


Figure 4.5: A realization of the proposed subband-dependent multivariate thresholding for redundant representation.

4.4.3 Computational Complexity

The computational complexity of the proposed multivariate SURE-LET algorithm can be evaluated in the same way as in Section 4.3.6. The main differences are that the number of linear parameters K is doubled and $(2M+1)^2 J$ partial derivatives must be computed. Taking into account these additional operations leads to an overall number of $\sim 2'600$ (resp. $\sim 15'700$) operations per pixel for denoising a 256×256 image with the UWT (resp. with the 12×12 BDCT), in a standard setting where $M = 2$.

4.5 Multichannel Estimator

In this section, we propose a vector/matrix extension of SURE-LET denoising in an arbitrary transform-domain. We use the same notations as the ones introduced in Section 3.4 of Chapter 3; i.e., the noisy multichannel image is denoted by

$$\mathbf{y} = [\mathbf{y}_1\ \mathbf{y}_2 \dots \mathbf{y}_N], \text{ where } \mathbf{y}_n = [y_{1,n}\ y_{2,n} \dots y_{C,n}]^{\mathrm{T}}$$

where C is the number of channels (e.g. $C = 3$ for color images).

Recall that, according to these notations, the MSE between an estimate

$$\hat{\mathbf{x}} = \mathbf{F}(\mathbf{y}) = \left(\mathbf{R}\boldsymbol{\Theta}(\underbrace{\mathbf{D}\mathbf{y}^{\mathrm{T}}}_{\mathbf{w}})\right)^{\mathrm{T}} \tag{4.19}$$

of the noise-free image $\mathbf{x}$ and the actual value of $\mathbf{x}$ is given by

$$\mathrm{MSE} = \frac{1}{CN}\mathrm{trace}\left\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^{\mathrm{T}}\right\}$$

4.5.1 Multichannel SURE

Under these conditions, Theorem 1 can be re-expressed as:

Corollary 6. *The following random variable*

$$\epsilon = \frac{1}{CN}\left(\mathrm{trace}\left\{(\mathbf{F}(\mathbf{y}) - \mathbf{y})(\mathbf{F}(\mathbf{y}) - \mathbf{y})^{\mathrm{T}} - N\boldsymbol{\Gamma}\right\} + 2\mathrm{div}\left\{\boldsymbol{\Gamma}\mathbf{F}(\mathbf{y})\right\}\right) \tag{4.20}$$

is an unbiased estimate of the expected MSE, i.e.

$$\mathscr{E}\left\{\epsilon\right\} = \frac{1}{CN}\mathscr{E}\left\{\mathrm{trace}\left\{(\mathbf{F}(\mathbf{y}) - \mathbf{x})(\mathbf{F}(\mathbf{y}) - \mathbf{x})^{\mathrm{T}}\right\}\right\}$$

where $\mathrm{div}\left\{\mathbf{F}(\mathbf{y})\right\} = \sum_{c=1}^{C}\sum_{n=1}^{N}\frac{\partial f_{c,n}(\mathbf{y})}{\partial y_{c,n}}$ *is a generalized divergence operator.*

The above result is a straightforward extension to the result proved in Theorem 1.

4.5.2 Multichannel LET

In the LET framework, $\mathbf{F}$ is described as a linear combination of elementary multichannel thresholding functions $\mathbf{F}_k$, i.e.

$$\begin{aligned}\mathbf{F}(\mathbf{y}) &= \sum_{k=1}^{K}\mathbf{a}_k^{\mathrm{T}}\underbrace{\left(\mathbf{R}\boldsymbol{\Theta}_k(\mathbf{D}\mathbf{y}^{\mathrm{T}})\right)^{\mathrm{T}}}_{\mathbf{F}_k(\mathbf{y})} \\ &= \underbrace{[\mathbf{a}_1^{\mathrm{T}}\ \mathbf{a}_2^{\mathrm{T}}\ \ldots\ \mathbf{a}_K^{\mathrm{T}}]}_{\mathbf{A}^{\mathrm{T}}}\begin{bmatrix}\mathbf{F}_1(\mathbf{y})\\ \mathbf{F}_2(\mathbf{y})\\ \vdots\\ \mathbf{F}_K(\mathbf{y})\end{bmatrix}\end{aligned} \tag{4.21}$$

where $\mathbf{A}$ is the $KC \times C$ matrix of unknown parameters.

We only consider here a (spatially) *pointwise* multichannel thresholding function, i.e.

$$\boldsymbol{\Theta}(\mathbf{w}) = \left[\boldsymbol{\theta}_l(\mathbf{w}_l)^{\mathrm{T}}\right]_{1\leq l\leq L} \text{ with } \boldsymbol{\theta}_l(\mathbf{w}_l) = \left[\theta_{l,c}(\mathbf{w}_l)\right]_{1\leq c\leq C} \tag{4.22}$$

In this case, the divergence term which appears in (4.20) is re-expressed as

$$\begin{aligned}
\operatorname{div}\{\boldsymbol{\Gamma}\mathbf{F}(\mathbf{y})\} &= \sum_{c=1}^{C}\sum_{n=1}^{N}\sum_{i=1}^{C}\gamma_{c,i}\frac{\partial f_{i,n}(\mathbf{y})}{\partial y_{c,n}} \\
&= \sum_{c=1}^{C}\sum_{i=1}^{C}\gamma_{c,i}\sum_{l=1}^{L}\frac{\partial\theta_{l,i}(\mathbf{w}_l)}{\partial w_{l,c}}\underbrace{\sum_{n=1}^{N}d_{l,n}r_{n,l}}_{[\mathbf{DR}]_{l,l}} \\
&= \sum_{c=1}^{C}\sum_{i=1}^{C}\gamma_{c,i}\,\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'_{i,c}(\mathbf{w}) \\
&= \operatorname{trace}\left\{\boldsymbol{\Gamma}\begin{bmatrix}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'_{1,1}(\mathbf{w}) & \cdots & \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'_{1,C}(\mathbf{w})\\ \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'_{C,1}(\mathbf{w}) & \cdots & \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'_{C,C}(\mathbf{w})\end{bmatrix}\right\}
\end{aligned} \tag{4.23}$$

where $\boldsymbol{\alpha} = \operatorname{diag}\{\mathbf{DR}\}$ and $\boldsymbol{\Theta}'_{i,j}(\mathbf{w}) = \left[\frac{\partial\theta_{l,i}(\mathbf{w}_l)}{\partial w_{l,j}}\right]_{1\leq l\leq L}$, for $1\leq i,j\leq C$.

Using the multichannel LET expression (4.21) and the above result (4.23) allows us to rewrite the unbiased estimate (4.20) of the MSE as

$$\epsilon = \frac{1}{CN}\operatorname{trace}\left\{\mathbf{A}^{\mathrm{T}}\mathbf{M}\mathbf{A} - 2\mathbf{C}^{\mathrm{T}}\mathbf{A}\right\} + \frac{1}{CN}\operatorname{trace}\left\{\mathbf{y}\mathbf{y}^{\mathrm{T}} - N\boldsymbol{\Gamma}\right\} \tag{4.24}$$

where $\mathbf{M} = [\mathbf{m}_{k,l}]_{1\leq k,l\leq K}$ and $\mathbf{C} = [\mathbf{c}_k]_{1\leq k\leq K}$, with:

$$\begin{cases}
\mathbf{m}_{k,l} = \mathbf{F}_k(\mathbf{y})\mathbf{F}_l(\mathbf{y})^{\mathrm{T}} \\
\mathbf{c}_k = \mathbf{F}_k(\mathbf{y})\mathbf{y}^{\mathrm{T}} - \boldsymbol{\Gamma}\begin{bmatrix}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'_{k;1,1}(\mathbf{w}) & \cdots & \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'_{k;1,C}(\mathbf{w})\\ \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'_{k;C,1}(\mathbf{w}) & \cdots & \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Theta}'_{k;C,C}(\mathbf{w})\end{bmatrix}
\end{cases} \tag{4.25}$$

The optimal parameters $\mathbf{A}_{\mathrm{opt}}$ of the linear expansion of thresholds (4.21) are the ones that minimize the unbiased MSE estimate given in (4.24), i.e.

$$\mathbf{A}_{\mathrm{opt}} = \mathbf{M}^{-1}\mathbf{C}$$

where the components of the $KC\times KC$ matrix $\mathbf{M}$ and of the $KC\times 1$ vector $\mathbf{C}$ are defined in (4.25).

4.5.3 Multichannel Thresholding

We propose a natural multichannel extension of the subband-adaptive grayscale denoiser introduced in Section 4.3.3, by taking into account the potentially strong interchannel similarities as follows:

$$\boldsymbol{\theta}_l(\mathbf{w}_l) \quad = \quad \mathbf{a}_1^{\mathrm{T}}\mathbf{w}_l + \gamma\left(\mathbf{w}_l^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\mathbf{w}_l\right)\mathbf{a}_2^{\mathrm{T}}\mathbf{w}_l \tag{4.26}$$

where $\gamma(x) = e^{-\left(\frac{x}{T}\right)^4}$ is used as the discriminator between large/small transformed coefficients. In our experiments, we found that $T = 9\sqrt{C}$ gave the best PSNR results. If $C = 1$, we obviously recover the grayscale thresholding function presented in Section 4.3.3.

4.6 Experiments

4.6.1 Transform-Domain vs Image-Domain Optimization

Before comparing the proposed general transform-domain *SURE-LET* approach with the best state-of-the-art algorithms, we demonstrate here that, in order to optimize the denoising process, it is essential to perform the minimization in the *image-domain*[4]. By contrast, an *independent wavelet subband* processing is suboptimal, often by a significant margin, even in a "tight" frame representation. This is because we usually do not have energy preservation between the denoised "tight" frame coefficients $\hat{\mathbf{w}}$ and the reconstructed image $\hat{\mathbf{x}} = \mathbf{R}\hat{\mathbf{w}}$: $\|\mathbf{R}\hat{\mathbf{w}}\| \neq \|\hat{\mathbf{w}}\|$. This is not in contradiction with the well-known energy conservation between the "tight" frame coefficients $\mathbf{w} = \mathbf{D}\mathbf{y}$ and the noisy image $\mathbf{y}$: $\|\mathbf{D}\mathbf{y}\| = \|\mathbf{y}\|$.

In Figure 4.6, we compare a classical transform-domain (TD) SURE-based optimization of our pointwise thresholding function (4.11) with the image-domain (ID) optimization based on Corollary 4 in the framework of the undecimated wavelet transform (considering two different wavelet filters *Haar* and *sym8*). We notice that the rigorous image domain optimization provides large improvements (up to +1 dB) over the independent in-band optimization when using the *Haar* filters. The same observation holds for the smoother *sym8* filters, but the gain is less important (around +0.5 dB). On average, the best results are obtained with an image-domain optimization of the undecimated *Haar* wavelet representation; this is not the case with non-redundant wavelet transforms, where the use of smoother wavelets (such as *sym8*) gives better results.

A closer examination of the "optimal" thresholding functions either obtained by the transform-domain or the image-domain SURE optimization indicates that the variation in PSNR performance may be related to the difference between the slopes of these functions around zero: the image-domain solution is actually much flatter, making it able to suppress small coefficients almost exactly.

[4]This fact has been also recently pointed out by Raphan *et. al.* in [117].

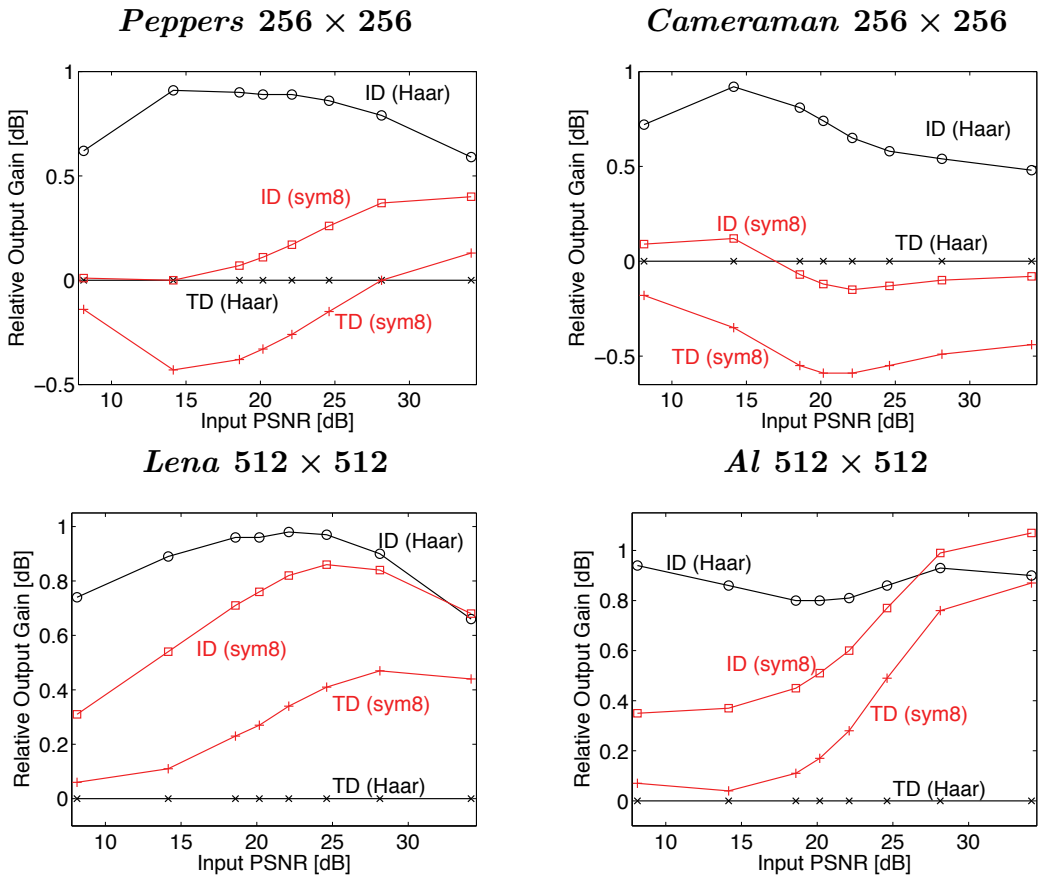


Figure 4.6: Comparison of the proposed *SURE-LET* denoising procedure (image-domain optimization) with a SURE-based denoising algorithm optimized in the wavelet domain when using the undecimated wavelet transform.

4.6.2 Periodic vs Symmetric Boundary Extensions

It is also worth quantifying the effects of particular boundary extensions. In Figure 4.7, we compare symmetric boundary extensions (rigorous SURE computation, as described in Section 4.3.4) with the periodic ones. As it can be observed, the symmetric boundary extension can lead to up to +0.5 dB of PSNR improvements over the periodic one.

4.6.3 Orthogonal vs Redundant Representation

In this section, we compare the pointwise *undecimated* (*Haar*) wavelet thresholding introduced in (4.11) and its multivariate extension (4.18), with the pointwise *orthonormal* (*sym8*) thresholding presented in (3.11) and its interscale multivariate extension described in Section 3.4.4. Periodic boundary extensions have been

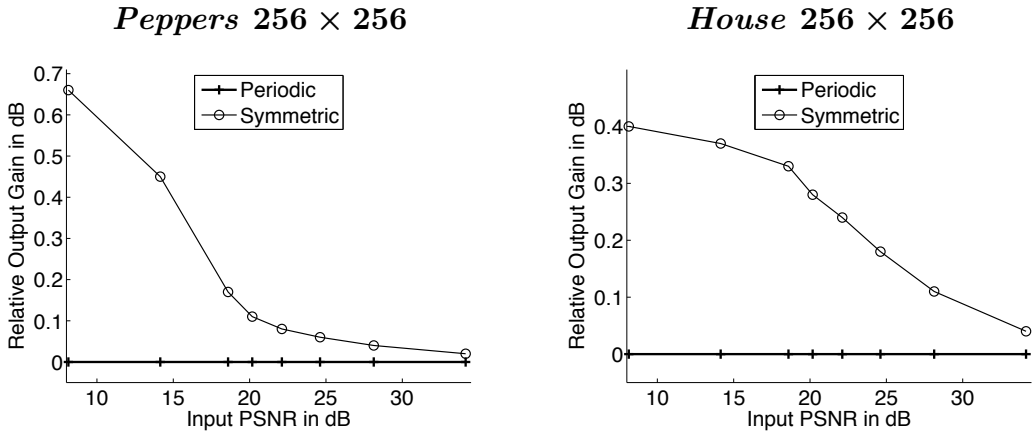


Figure 4.7: Influence of the boundary extensions when using the undecimated *Haar* wavelet transform.

considered for all methods and the same number of decomposition levels has been performed. Figure 4.8 summarizes the following observations:

- A pointwise thresholding applied in a *redundant* wavelet representation significantly outperforms (up to +2.5 dB) a similar complexity pointwise thresholding applied in an *orthonormal* wavelet decomposition.
- A *multivariate* thresholding applied in an *orthonormal* wavelet decomposition is generally outperformed ($0.5-1$ dB loss) by a *pointwise* thresholding applied in a *redundant* wavelet representation.
- The PSNR improvements brought by the application of a more sophisticated (i.e. multivariate) thresholding are less significant in redundant representation than in orthonormal wavelet decomposition.

4.6.4 UWT vs Overcomplete BDCT

In Figure 4.9, we compare the PSNRs obtained by the application of the pointwise SURE-LET thresholding in the Haar undecimated wavelet transform and in an overcomplete 12×12 BDCT representation. On the one hand, we observe that the Haar UWT is a better transform for simple, approximately piecewise-smooth images such as *Peppers* and *Cameraman*. On the other hand, the BDCT is more efficient at yielding a sparse representation of the various textures present in images like *Lena* and *Barbara*.

In order to get the best out of several transforms, we propose to make the LET span several transformed domains with complementary properties: SURE is then

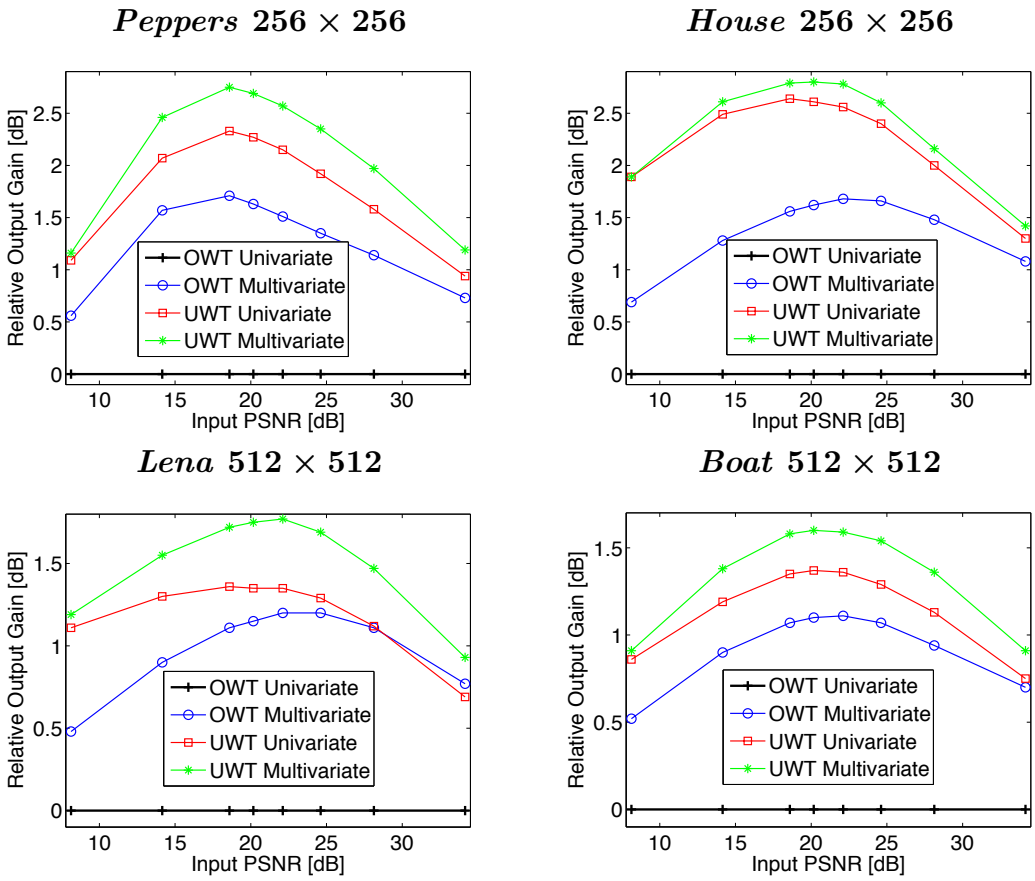


Figure 4.8: Comparison between SURE-LET algorithms applied in an *orthonormal* wavelet decomposition and SURE-LET algorithms applied in a *redundant* wavelet representation.

used to globally optimize the weights of this linear combination of processed subbands. In this case, the union of several transforms can be interpreted as an overcomplete dictionary of bases which can sparsely represent a wide class of natural images. As expected, this combined and jointly optimized UWT/BDCT SURE-LET solution gives better results than either the UWT or the BDCT SURE-LET; it is also more efficient than any simple convex combinations of both.

Note that the idea of combining several complementary transforms was also exploited in the context of image denoising by Starck *et. al.* in [118]. The use of an overcomplete dictionary, either fixed in advance (as in our case) or trained, is at the core of the *K-SVD*-based denoising algorithm of Elad *et. al.* [34].

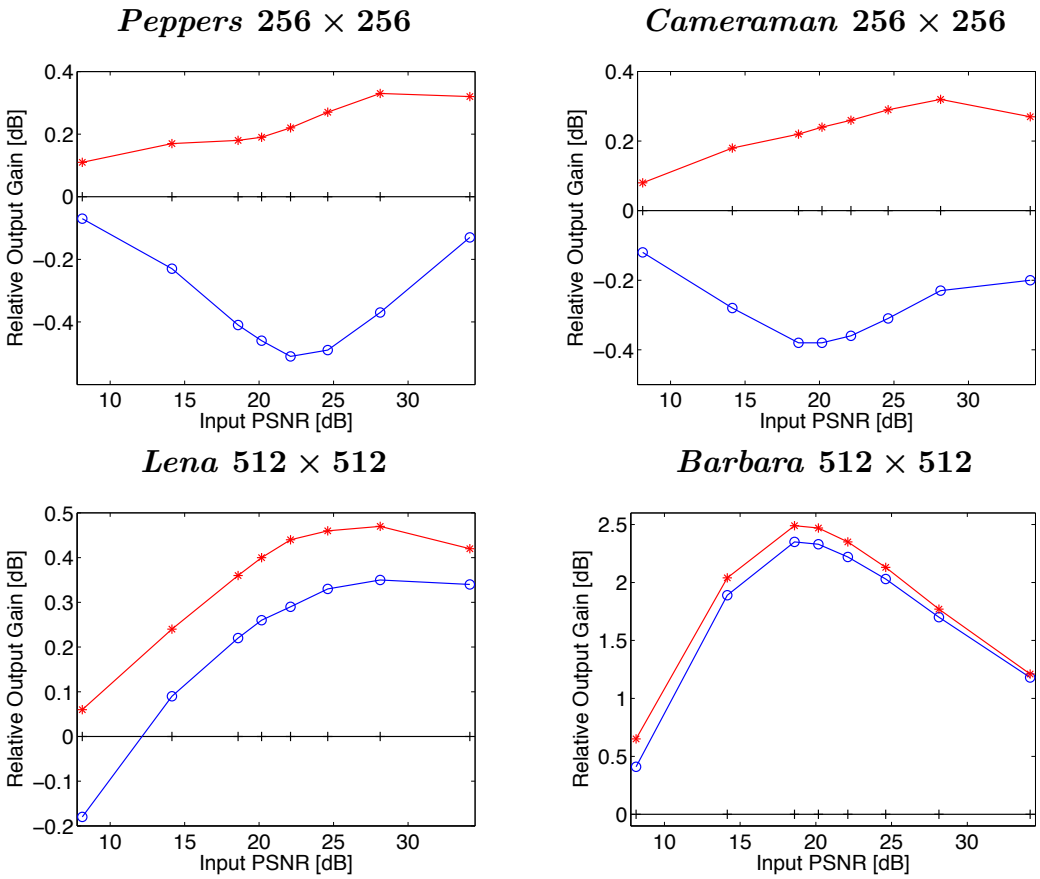


Figure 4.9: Comparison between the pointwise SURE-LET thresholding (4.11) applied in the Haar undecimated wavelet transform ("+" markers, benchmark), in an overcomplete 12×12 BDCT representation ("$\circ$" markers) and in an UWT/BDCT dictionary ("$*$" markers).

4.6.5 Comparisons with Multiscale Image Denoising Methods

In this section, we compare the proposed pointwise SURE-LET approach with some of the best state-of-the-art *redundant* multiresolution-based techniques. We have retained the same methods as in Section 3.7.1 of Chapter 3 but applied in a redundant representation; i.e., *BiShrink* [84] applied with a dual-tree complex wavelet transform (DTCWT), *ProbShrink* [82] applied in an undecimated wavelet representation using Daubechies *symlets* filters with eight vanishing moments (*sym8*), and *BLS-GSM* [81] applied in a full steerable (eight orientations per scale) pyramid (FSP). We have followed the same protocol as in Section 3.7.1, except that symmetric boundary extensions have been applied for all methods.

Table 4.1 reports the PSNR results obtained by the various denoising methods,

Table 4.1: Comparison of some of the most efficient redundant multiscale denoising methods.

σ	5	10	15	20	25	30	50	100
Input PSNR	34.15	28.13	24.61	22.11	20.17	18.59	14.15	8.13
Method	**Peppers 256 × 256**							
BiShrink	37.17	33.37	31.27	29.80	28.67	27.76	25.27	22.10
ProbShrink	37.33	33.50	31.34	29.81	28.65	27.72	25.17	21.97
BLS-GSM	37.34	33.78	31.74	30.31	29.21	28.31	25.85	22.58
UWT SURE-LET (4.11)	37.63	34.00	31.96	30.52	29.39	28.47	25.96	22.66
BDCT SURE-LET (4.11)	37.50	33.63	31.47	30.01	28.93	28.06	25.73	22.60
Dictionary SURE-LET (4.11)	**37.95**	**34.33**	**32.23**	**30.74**	**29.58**	**28.65**	**26.13**	**22.78**
Method	**House 256 × 256**							
BiShrink	38.36	34.73	32.92	31.66	30.67	29.85	27.54	24.46
ProbShrink	38.27	34.80	32.98	31.66	30.63	29.77	27.41	24.20
BLS-GSM	38.71	35.39	33.66	32.40	31.39	30.54	28.16	24.84
UWT SURE-LET (4.11)	38.72	35.53	33.84	32.64	31.70	30.93	28.59	25.21
BDCT SURE-LET (4.11)	39.28	35.74	33.80	32.48	31.48	30.65	28.36	25.05
Dictionary SURE-LET (4.11)	**39.40**	**36.03**	**34.20**	**32.94**	**31.95**	**31.12**	**28.69**	**25.21**
Method	**Barbara 512 × 512**							
BiShrink	37.35	33.51	31.37	29.87	28.71	27.78	25.29	22.46
ProbShrink	37.39	33.49	31.24	29.60	28.33	27.30	24.54	22.00
BLS-GSM	37.79	34.02	31.83	30.29	29.09	28.12	25.42	22.52
UWT SURE-LET (4.11)	36.98	32.65	30.16	28.45	27.18	26.23	24.14	22.27
BDCT SURE-LET (4.11)	38.16	34.35	32.19	30.67	29.51	28.58	26.03	22.68
Dictionary SURE-LET (4.11)	**38.19**	**34.42**	**32.29**	**30.80**	**29.65**	**28.72**	**26.18**	**22.92**
Method	**Boat 512 × 512**							
BiShrink	36.72	33.18	31.31	29.99	28.97	28.16	25.99	23.34
ProbShrink	36.69	33.29	31.35	29.98	28.91	28.07	25.85	23.19
BLS-GSM	36.98	33.58	31.70	30.37	29.36	28.54	26.35	23.65
UWT SURE-LET (4.11)	37.13	33.54	31.58	30.23	29.22	28.41	26.23	23.63
BDCT SURE-LET (4.11)	37.21	33.52	31.54	30.18	29.14	28.31	26.14	23.52
Dictionary SURE-LET (4.11)	**37.38**	**33.78**	**31.83**	**30.48**	**29.44**	**28.60**	**26.36**	**23.69**
Method	**Al 512 × 512**							
BiShrink	38.72	35.34	33.52	32.26	31.28	30.47	28.20	25.06
ProbShrink	38.79	35.35	33.49	32.14	31.02	30.26	28.06	24.93
BLS-GSM	38.99	35.59	33.83	32.62	31.69	30.92	28.73	25.67
UWT SURE-LET (4.11)	38.88	35.43	33.61	32.37	31.43	30.67	28.60	25.73
BDCT SURE-LET (4.11)	39.20	35.74	33.91	32.65	31.69	30.90	28.69	25.65
Dictionary SURE-LET (4.11)	**39.27**	**35.83**	**34.01**	**32.77**	**31.81**	**31.04**	**28.87**	**25.86**
Method	**Goldhill 512 × 512**							
BiShrink	36.78	33.12	31.24	30.00	29.09	28.38	26.53	24.21
ProbShrink	36.57	32.97	31.07	29.83	28.93	28.23	26.42	24.21
BLS-GSM	37.00	33.38	31.52	30.30	29.40	28.69	26.83	24.52
UWT SURE-LET (4.11)	36.85	33.21	31.37	30.17	29.29	28.61	26.83	24.70
BDCT SURE-LET (4.11)	37.05	33.39	31.52	30.29	29.38	28.67	26.83	24.57
Dictionary SURE-LET (4.11)	**37.19**	**33.57**	**31.70**	**30.46**	**29.55**	**28.83**	**26.95**	**24.72**

Note: The output PSNRs have been averaged over ten noise realizations.

the best results being shown in boldface. In most cases, the proposed *pointwise* SURE-LET thresholding applied in the *Haar* UWT achieves higher PSNRs than both the *multivariate BiShrink* and *ProbShrink* (average gain of 0.3 dB). It also matches the performance of the best multiresolution-based technique (*BLS-GSM*). As emphasized in Section 4.6.4, it is more advisable to apply the SURE-LET strategy to an overcomplete BDCT to achieve more competitive results on some images such as *Barbara*. In all cases, the SURE-LET thresholding optimized in a UWT/BDCT dictionary achieves the highest PSNR results.

In Figure 4.10, we show some visual comparisons between the various multiscale denoising algorithms applied on *Barbara*. As observed, our dictionary-based SURE-LET algorithm can efficiently restore the various textures present in this particular image. This observation is confirmed by the SSIM visual quality measure.

4.6.6 Comparisons with Other State-of-the-art Image Denoising Methods

In the previous section, we have compared our *pointwise* SURE-LET denoising strategy with the most efficient state-of-the-art *multiresolution-based* techniques. We propose now to compare our *multivariate* approach with some of the most successful alternative denoising solutions:

- The non-local means (*NL-means*) algorithm of Buades *et. al.* [7]: this algorithm belongs to the class of patch-based approaches (see Section 1.3.2) and involves three important parameters that need to be fixed. The degree of smoothing has been set to 0.7σ, a value that has been recently shown to give near optimal performances [25]. The size of the search window has been set to 21×21, which constitutes a good trade-off between PSNR performances and computation time. For a fair comparison, we have optimized (in the MMSE sense) the size of the similarity window (from 3×3 to 9×9).

- The total-variation (TV) minimization by Rudin *et. al.* [28]: the principle of this denoising algorithm has been sketched in Section 1.3.3. In practice, we have used the fast implementation proposed by Chambolle[5] [119]. In order to get the highest PSNR performances, we have optimized (in the MMSE sense) the value of the regularization parameter in a finite set of 50 values logarithmically ranging from 10^{-4} to 1.

- The *K-SVD* algorithm of Elad *et. al.* [34]: at the core of their denoising approach lies the notion of *Sparseland* model, which assumes that a broad class of images can be well represented by a linear combination of few atoms from an overcomplete dictionary. The latter can be either fixed in advance (e.g.

[5] Matlab implementation available at:
`http://www.math.ucla.edu/~getreuer/matlabimaging.html#topics`

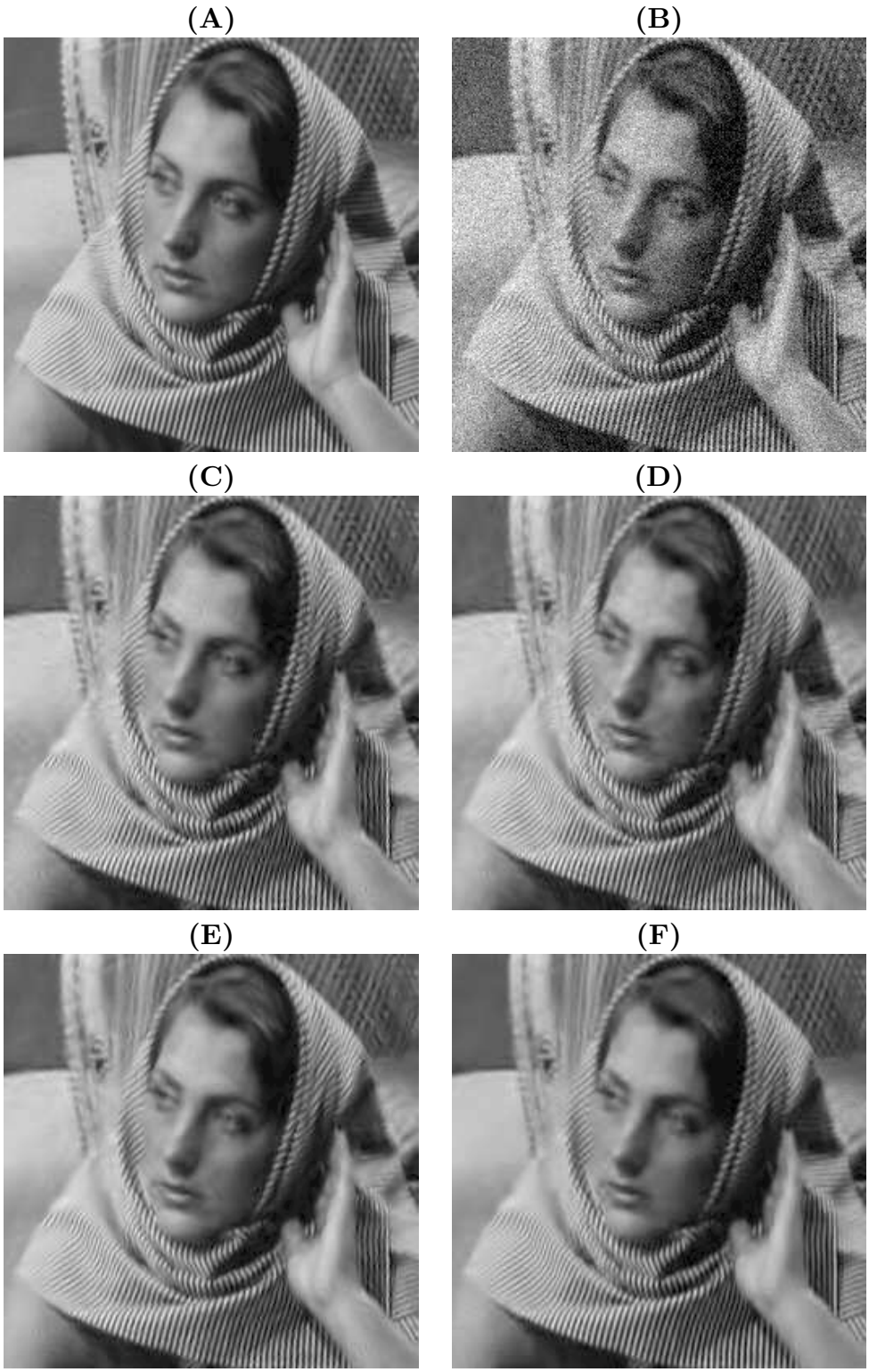


Figure 4.10: (**A**) Part of the noise-free *Barbara*. (**B**) Noisy version of it: SSIM = 0.48. (**C**) Denoised by *UWT-ProbShrink*: SSIM = 0.85. (**D**) Denoised by *DTCWT-BiShrink*: SSIM = 0.86. (**E**) Denoised by *FSP-BLS-GSM*: SSIM = 0.87. (**E**) Denoised by the proposed pointwise *SURE-LET*, optimized in an UWT/BDCT dictionary: SSIM = 0.88.

an overcomplete BDCT), trained on a set of image patches taken from good quality images or trained on the noisy image itself. If the training option is retained, then the redundant dictionary is adaptively build using the so-called *K-SVD* algorithm [120]. The denoised image is the minimizer of a regularized cost functional involving a global data-fidelity term and local image priors which ensure that each patch of the denoised image has a sparse representation in the considered dictionary.

- The sparse 3D transform-domain collaborative filtering (*BM3D*[6]) by Dabov *et. al.* [121]: this solution is a sophisticated two-pass algorithm which can be briefly summarized as follows. In the first pass, a block-matching procedure is applied to each square block of the noisy image. In the neighborhood of each of these reference blocks, a search for similar blocks is performed. The best matching blocks and their corresponding reference block are grouped together to yield a 3D stack of similar blocks. A 3D transform-domain thresholding is then applied to the stack. An estimate of each pixel is finally obtained by aggregation (i.e. weighted combination) of the block-wise estimates that overlap with the current pixel. In the second pass, the same procedure is applied, except that the block-matching is performed on the denoised image (as estimated by the first step) and the 3D transform-domain thresholding is replaced by an approximated *oracle* pointwise Wiener filter which makes use of the denoised image as an estimation of the underlying noise-free image. Although many parameters involved in this algorithm are empirically chosen, the *BM3D* currently achieves the highest PSNR performances and exhibits the fewest artifacts for most images and for a wide range of noise levels.

In Figure 4.11, we display some PSNR comparisons between these algorithms and the proposed multivariate SURE-LET thresholding (4.18), globally optimized in a Haar UWT/12×12 BDCT dictionary. As observed, the proposed solution outperforms both the *NL-means* and the *TV* algorithms (around $+1$ dB). We consistently obtain higher PSNRs ($+0.3$ dB on average) than the *K-SVD* approach, which tends to further degrade as the noise level increases. Yet, we generally obtain slightly lower PSNRs (up to 0.5 dB) than the *BM3D*. We believe that the consideration of a more sophisticated/heterogenous dictionary (containing basis elements with higher directional selectivity, such as the *curevelets* [52] or the full steerable pyramid [47]) in a SURE-LET framework could fill this little gap.

In Figure 4.12, we show the denoising result of each of the algorithms described in this section. The computed SSIM values confirm the PSNR ranking of the various methods: the proposed solution holds its own against the high-quality *BM3D*.

From a computational point of view, a UWT/BDCT dictionary-based multivariate ($M = 2$) SURE-LET algorithm is much more time-consuming than a simple

[6]Matlab implementation available at:
`http://www.cs.tut.fi/~foi/GCF-BM3D/#ref_software`

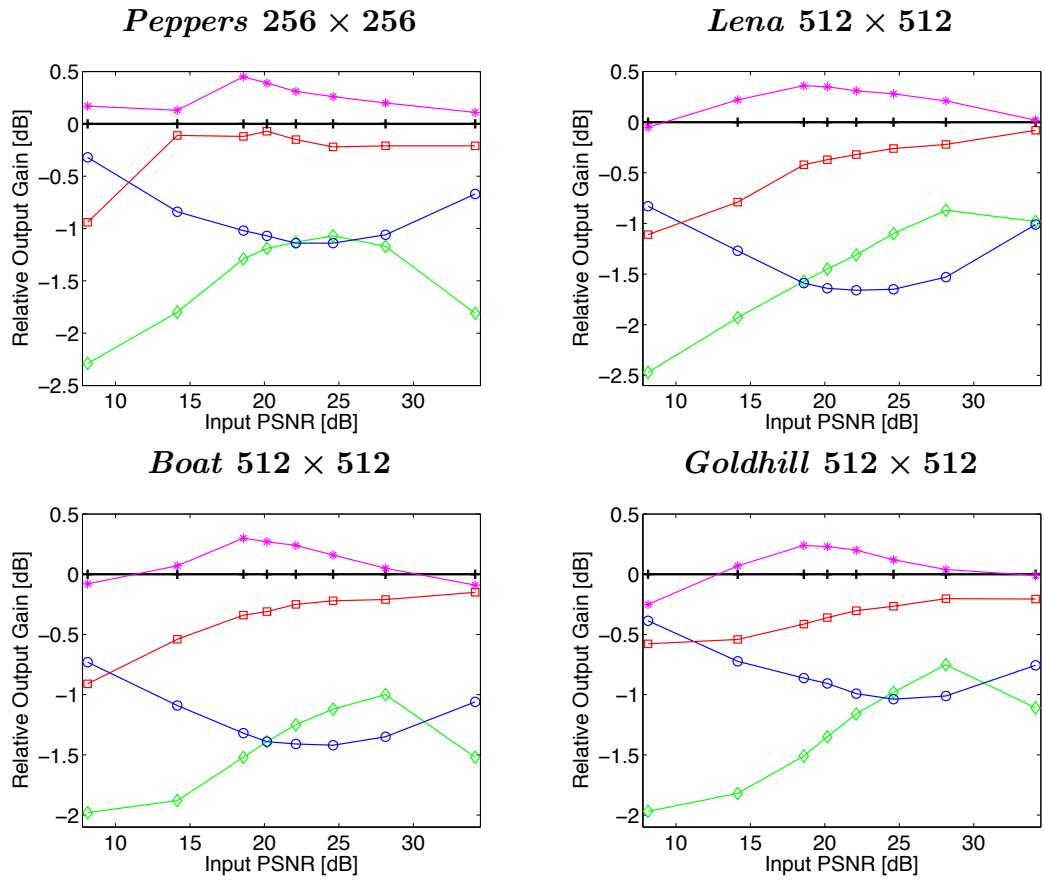


Figure 4.11: Comparison between the proposed multivariate *SURE-LET* algorithm ("+" markers, benchmark) optimized in an UWT/BDCT dictionary and some successful denoising algorithms: *NL-means* [7] ("◊" markers), *TV* [28] ("○" markers), *K-SVD* [34] ("□" markers) and *BM3D* [121] ("∗" markers).

pointwise SURE-LET thresholding applied in either the UWT or in the BDCT representation, because it requires around $20'400$ operations per pixel. Yet, comparing to the best state-of-the-art denoising methods, it still constitutes a competitive solution. Indeed, the *K-SVD* [34] and the *BM3D* [121] require respectively $\sim 163'840$ and at most $31'600$ operations per pixel[7]. The proposed algorithm is thus approximately one order of magnitude faster than the *K-SVD*, while having a similar computational complexity to that of the *BM3D*.

[7]These values have been computed based on the computational complexity formulas given by the authors in their respective publication, and according to the parameters setting used in their implementation.

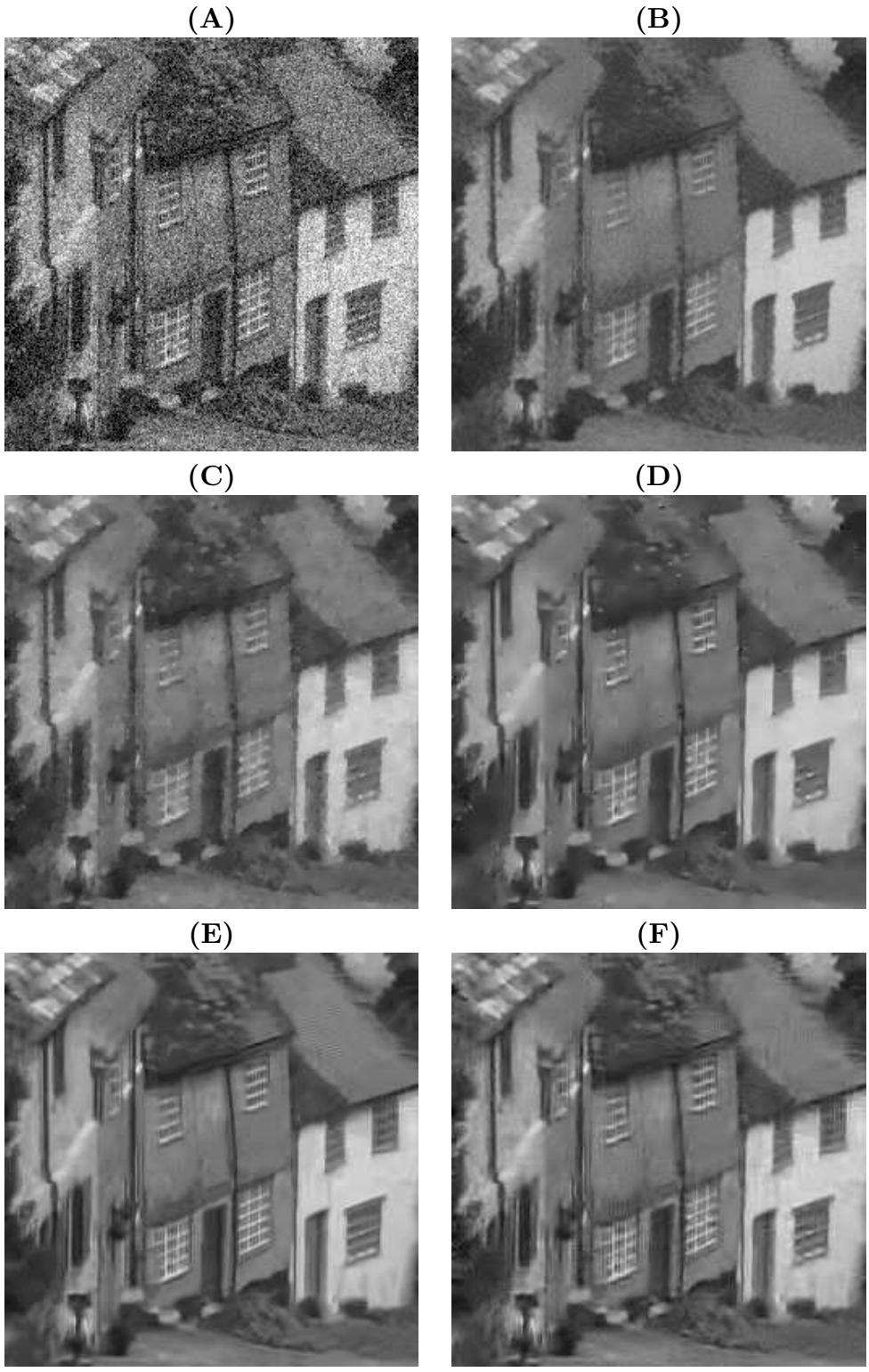


Figure 4.12: (**A**) Part of the noisy *Goldhill*: SSIM = 0.48. (**B**) Denoised by *NL-means*: SSIM = 0.66. (**C**) Denoised by *TV*: SSIM = 0.70. (**D**) Denoised by *K-SVD*: SSIM = 0.71. (**E**) Denoised by *BM3D*: SSIM = 0.75. (**E**) Denoised by the proposed multivariate *SURE-LET* optimized in an UWT/BDCT dictionary: SSIM = 0.74.

4.6.7 Comparisons with State-of-the-Art Color Image Denoising Methods

We have applied our multichannel SURE-LET thresholding for denoising color images using the *Haar* UWT with *symmetric* boundaries handling. A new thresholding function (4.26) was applied inside each of the wavelet subbands, while keeping the lowpass residual unchanged. The experiments have been executed on the set of standard RGB images shown in Figure 3.16 using the same protocol as the one described in Section 3.7.2.

Table 4.2: Comparison of some color image denoising algorithms (same noise level in each RGB channel)

$\sigma_R = \sigma_G = \sigma_B$	5	10	15	20	25	30	50	100
Input PSNR [dB]	34.15	28.13	24.61	22.11	20.17	18.59	14.15	8.13
Method	***Image 1* 256 × 256**							
ProbShrink-MB [82]	37.99	34.67	32.97	31.81	30.93	30.18	27.93	24.70
SA-DCT [65]	**38.60**	**35.64**	**34.07**	**32.97**	**32.10**	**31.39**	**29.18**	**25.90**
UWT SURE-LET (4.26)	38.51	35.44	33.88	32.80	31.95	31.25	29.11	25.75
Method	***Image 2* 256 × 256**							
ProbShrink-MB [82]	35.39	31.23	29.32	28.14	27.25	26.56	24.67	22.20
SA-DCT [65]	35.50	31.26	29.44	28.40	27.64	27.04	**25.31**	**22.85**
UWT SURE-LET (4.26)	**35.57**	**31.53**	**29.70**	**28.55**	**27.71**	**27.05**	25.23	22.72
Method	***Image 3* 512 × 512**							
ProbShrink-MB [82]	37.44	34.35	32.74	31.60	30.76	30.03	28.03	25.43
SA-DCT [65]	37.69	34.97	**33.60**	**32.63**	**31.86**	**31.21**	**29.25**	26.47
UWT SURE-LET (4.26)	**38.03**	**35.05**	33.55	32.51	31.70	31.03	29.14	**26.53**
Method	***Image 4* 512 × 512**							
ProbShrink-MB [82]	36.47	33.46	32.06	31.11	30.36	29.72	27.81	25.00
SA-DCT [65]	36.78	33.68	32.39	31.55	**30.88**	**30.30**	28.50	25.76
UWT SURE-LET (4.26)	**36.89**	**33.86**	**32.49**	**31.58**	30.87	30.28	**28.52**	**25.81**
Method	***Image 5* 512 × 512**							
ProbShrink-MB [82]	34.85	30.06	27.63	26.13	25.04	24.18	22.01	19.94
SA-DCT [65]	**35.22**	30.61	28.32	26.88	**25.85**	**25.06**	**23.01**	20.52
UWT SURE-LET (4.26)	35.20	**30.67**	**28.40**	**26.92**	25.84	25.00	22.88	**20.68**
Method	***Image 6* 512 × 512**							
ProbShrink-MB [82]	38.82	35.31	33.42	32.12	31.15	30.32	27.86	24.59
SA-DCT [65]	39.46	36.37	34.64	**33.39**	**32.41**	**31.59**	29.22	26.04
UWT SURE-LET (4.26)	**39.57**	**36.44**	**34.66**	**33.39**	32.40	**31.59**	**29.24**	**26.05**

Notes: The output PSNRs have been averaged over ten noise realizations.

In Table 4.2 we compare our PSNR results with those obtained by running two other state-of-the-art color image denoising algorithms:

- Pižurica *et al.* *ProbShrink-MB* using the same transform (UWT Haar) and boundary conditions (symmetric) as ours.

- Foi *et al.* pointwise *SA-DCT*[8] [65], which is the application of their grayscale shape-adaptive DCT denoiser in the opponent color space, but using the adaptive neighborhoods defined in the luminance channel for all channels.

We notice that we obtain a significant gain (about +1 dB) over the *ProbShrink-MB*, and similar results to the pointwise *SA-DCT*. Moreover, our denoised images contain very few color artifacts, and represent a good trade-off between noise removal and preservation of small details (see Figure 4.13).

From a computational point of view, the execution of our un-optimized Matlab implementation lasts around 3s for 256×256 RGB images, which is slightly faster than the two other algorithms that make use of pre-compiled codes.

4.7 Summary

In this chapter, we have generalized the SURE-LET approach to arbitrary transform-domain denoising. The specificity of SURE-LET for redundant or non-orthonormal transforms lies in the fact that the MSE/SURE minimization is performed *in the image domain*. While it is true that, due to some Parseval-like MSE conservation, image domain MSE/SURE minimization is equivalent to separate in-band MSE/SURE minimization whenever the analysis transformation is (non-redundant) orthonormal, this is grossly wrong as soon as the transformation is, either *redundant* (even when it is a "tight frame") or *non-orthonormal*. In our experiments, we have pointed out that a rigorous image-domain SURE-based parameters optimization brings a non-negligible PSNR improvement over the conventional independent subband optimization. Note that the computational efficiency of this image-domain optimization is only possible when considering a *linear* parameterization of the denoising process; this naturally makes prohibitive the use of sophisticated wavelet estimators that require the optimization of several non-linear parameters.

We have shown that a further development of the divergence of the denoising process which appears in the SURE expression reveals the diagonal of the product between the decomposition and the reconstruction matrices. Provided that these matrices are explicitly given, this divergence term can be computed analytically. When this is not the case, we have described a simple numerical procedure to compute a good approximation of this transform-dependent term.

Pointwise, multivariate and multichannel thresholding rules have been then proposed. Their application in the Haar UWT and in an overcomplete BDCT representation have been experimented, leading to the following observation: the choice of transform can have a significant influence on the denoising quality. To obtain the best performance out of several transformations, we have thus proposed to apply the SURE-LET strategy to an heterogenous dictionary of overcomplete transforms. The

[8] Matlab code available online at `http://www.cs.tut.fi/\~foi/SA-DCT/\#ref_software`

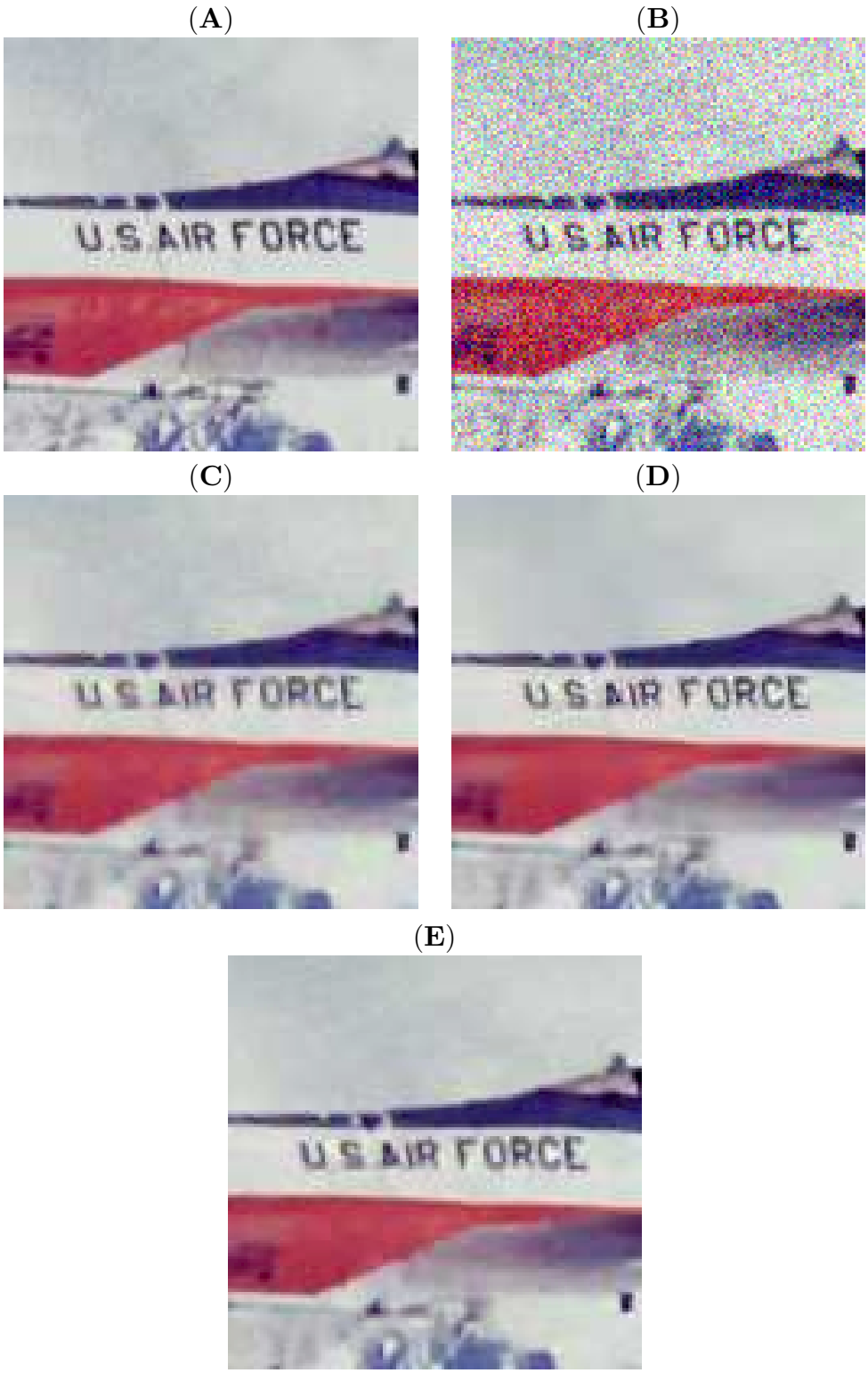


Figure 4.13: (**A**) Part of the noise-free *Image 6*. (**B**) Noisy version of it: PSNR = 18.59 dB. (**C**) Denoised by *ProbShrink-MB*: PSNR = 30.32 dB. (**D**) Denoised by the pointwise *SA-DCT*: PSNR = 31.60 dB. (**E**) Denoised by the proposed method: PSNR = 31.61 dB.

resulting dictionary-based SURE-LET has been finally shown to favorably compare with the state-of-the-art denoising algorithms.

Chapter 5

SURE-LET for Fast Video Denoising

5.1 Motivations

Despite the recent advances in video sequence acquisition and transmission, denoising still remains an essential step before performing higher level tasks, such as coding, compression, object tracking or pattern recognition. Since the origins of the degradations are numerous and diverse (imperfection of the CCD detectors, electronic instabilities, thermal fluctuations, ...), the overall noise contribution is often modeled as an additive (usually Gaussian) white process, independent from the original uncorrupted image sequence [122].

The huge amount of correlations present in every video sequences has quite early led the researchers to develop combined spatio-temporal denoising algorithms, instead of sequentially applying available 2D tools. The emergence of new multiresolution tools such as the wavelet transform [38, 39] then gave an alternative to the standard noise reduction filters that were used for video denoising [123–126]. Now, the transform-domain techniques in general, and the wavelet-based in particular [127–133], have been shown to outperform these spatio-temporal linear and even non-linear filtering.

In this chapter[1], we stay within this scope of wavelet-domain video denoising techniques. More precisely, and contrary to most of the existing techniques [127–130, 133], we consider an *orthonormal* wavelet transform rather than redundant representations, because of its appealing properties (energy and noise statistics preservation), its lower computational complexity and memory usage. To take into account the strong temporal correlations between adjacent fames, we work out a

[1] This chapter is based on our submitted paper [134].

multiframe wavelet thresholding based on the multichannel SURE-LET strategy introduced in Section 3.4. To increase the correlations between adjacent frames, we compensate for interframe motion using a global motion compensation followed by a *selective* block-matching procedure. The selectivity is obtained by first performing a coarse interframe motion detection and then only matching those blocks inside which a *significant* motion occurred. Thanks to its selectivity, the proposed block-matching has a negligible influence on the interframe noise covariance matrix. This latter point is crucial for the efficiency of our SURE-LET algorithm. Instead, standard block-matching [135] would make it difficult to track the interframe noise statistics.

5.2 Multiframe SURE-LET

In this section, we use the same formalism as in the multichannel SURE-LET denoising developed in Section 3.4. We denote an original (unknown) video sequence of T frames containing N pixels by:

$$\mathbf{v} = [\mathbf{v}_1\ \mathbf{v}_2\ \dots\ \mathbf{v}_N], \text{ with } \mathbf{v}_n = [v_{1,n}\ v_{2,n}\ \dots\ v_{T,n}]^{\mathrm{T}} \tag{5.1}$$

We also define a unitary $T \times 1$ vector $\mathbf{e}_t$ such that $\mathbf{e}_t^{\mathrm{T}}\mathbf{v}_n = v_{t,n}$, and we assume that the observed noisy video sequence is given by $\mathbf{u} = \mathbf{v} + \mathbf{n}$, where $\mathbf{n}$ is an additive white Gaussian noise independent of $\mathbf{v}$, with known $T \times T$ interframe covariance matrix $\mathbf{\Gamma}$.

In an *orthonormal* wavelet representation, the observation model is preserved in the transformed domain, as well as the interframe noise covariance matrix $\mathbf{\Gamma}$. Therefore, each noisy wavelet coefficient $\mathbf{y}_n^j \in \mathbb{R}^T$, $j = 1 \dots J$, $n = 1 \dots N^j$ is given by

$$\mathbf{y}_n^j = \mathbf{x}_n^j + \mathbf{b}_n^j, \text{ where } \mathbf{b}_n^j \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}) \tag{5.2}$$

5.2.1 Multiframe SURE

Denoting by $\hat{\mathbf{v}}$ an estimate of the noise-free video $\mathbf{v}$, we can define the global MSE as:

$$\begin{aligned} \mathrm{MSE} &= \frac{1}{NT}\sum_{t=1}^{T}\underbrace{\sum_{n=1}^{N}\mathbf{e}_t^{\mathrm{T}}(\hat{\mathbf{v}}_n - \mathbf{v}_n)(\hat{\mathbf{v}}_n - \mathbf{v}_n)^{\mathrm{T}}\mathbf{e}_t}_{N\times\mathrm{MSE}_t} \\ &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{J}\underbrace{\sum_{n=1}^{N^j}\mathbf{e}_t^{\mathrm{T}}(\hat{\mathbf{x}}_n^j - \mathbf{x}_n^j)(\hat{\mathbf{x}}_n^j - \mathbf{x}_n^j)^{\mathrm{T}}\mathbf{e}_t}_{N^j\times\mathrm{MSE}_t^j} \end{aligned} \tag{5.3}$$

where $\mathbf{e}_t^{\mathrm{T}}\hat{\mathbf{x}}_n^j = \theta_t^j(\mathbf{y}_n^j, \mathbf{p}_n^j)$ is the n^{th} pixel of the j^{th} wavelet subband of the denoised frame t. It is obtained by thresholding the n^{th} pixel of the j^{th} wavelet subband of the noisy frame t, taking into account (some of) its neighboring frames. From now on, we will drop the subband superscript "j" and the time frame indication "t" for the sake of clarity, when no ambiguities arise.

Considering this multiframe processing $\theta : \mathbb{R}^T \times \mathbb{R}^T \to \mathbb{R}$, the MSE of any wavelet subband j of any frame t is estimated *without bias* by

$$\epsilon = \frac{1}{N}\sum_{n=1}^{N}\Big[\big(\theta(\mathbf{y}_n, \mathbf{p}_n) - \mathbf{e}_t^{\mathrm{T}}\mathbf{y}_n\big)^2 + 2\mathbf{e}_t^{\mathrm{T}}\mathbf{\Gamma}\nabla_1\theta(\mathbf{y}_n, \mathbf{p}_n) - N\mathbf{e}_t^T\mathbf{\Gamma}\mathbf{e}_t\Big] \tag{5.4}$$

in accordance with the general SURE methodology developed so far.

Recall that ∇_1 stands for the gradient operator relatively to the *first* variable of the function θ, i.e. $\mathbf{y}_n$ (see Section 3.4). If $\mathbf{y}_n$ is considered as a vector of neighboring wavelet coefficients (instead of a multiframe wavelet coefficient), the above result (5.4) can be interpreted as a multivariate SURE, similar to the one used in Section 3.4.4 to optimize a multivariate thresholding.

5.2.2 Multiframe LET

The thresholding function is specified by the following linear combination of K basic thresholding functions

$$\theta(\mathbf{y}_n, \tilde{\mathbf{y}}_n) \;=\; \underbrace{[\mathbf{a}_1^{\mathrm{T}}\ \mathbf{a}_2^{\mathrm{T}}\ldots\mathbf{a}_K^{\mathrm{T}}]}_{\mathbf{a}^{\mathrm{T}}} \underbrace{\begin{bmatrix}\boldsymbol{\theta}_1(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\\ \boldsymbol{\theta}_2(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\\ \vdots \\ \boldsymbol{\theta}_K(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\end{bmatrix}}_{\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)} \tag{5.5}$$

where $\mathbf{a}$ and $\boldsymbol{\theta}$ are both $KT \times 1$ vectors.

Thanks to this linear parameterization, the optimal —in the minimum ϵ sense— parameters of (5.5) are the solution of the following *linear* system of equations:

$$\mathbf{a}_{\text{opt}} \;=\; \mathbf{M}^{-1}\mathbf{C} \tag{5.6}$$

where:
$$\begin{cases} \mathbf{M} &= \displaystyle\sum_{n=1}^{N}\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)^{\mathrm{T}} \\ \mathbf{C} &= \displaystyle\sum_{n=1}^{N}\Big(\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\mathbf{y}_n^{\mathrm{T}} - \big(\nabla_1\boldsymbol{\theta}(\mathbf{y}_n, \tilde{\mathbf{y}}_n)\big)^{\mathrm{T}}\mathbf{\Gamma}\Big)\mathbf{e}_t \end{cases}$$

5.3 A Fast Video Denoising Algorithm

We propose to de noise the video *frame by frame*, by considering a sliding temporal window of τ (odd) neighboring frames centered around the current frame. For instance, the denoising of the *reference* frame t involves frames $t-(\tau-1)/2$ to $t+(\tau-1)/2$.

The various steps of the proposed algorithm (Fig. 5.1) are the following: We first align all the neighboring frames (*global* registration) and compensate for their (*local*) motion, w.r.t. the frame t. Then, this reference frame is processed in the wavelet domain, using thresholds based on the values of the wavelet coefficients of the aligned neighboring frames, and on their own coarser-scale coefficients (multiframe interscale SURE-LET thresholding). Finally, an inverse wavelet transform is performed on the denoised coefficients of this reference frame. These steps are detailed in Sections 5.3.1, 5.3.2 and 5.3.3.

5.3.1 Global Motion Compensation

As a global motion model, we can simply consider the translations due to camera motions (pan/tilt). The optimal integer shift s_{opt} required to register a given frame $t+\Delta t$ with respect to the reference frame t, is the index of the maximum of the cross-correlation function between the two frames [133, 136], i.e.

$$s_{\text{opt}} \quad = \quad \text{argmax}_s \, \mathcal{F}^{-1}\left\{U_t(\cdot)U^*_{t+\Delta t}(\cdot)\right\}(s) \tag{5.7}$$

where $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse discrete Fourier transform and $U_t(\omega)$, $U_{t+\Delta t}(\omega)$ are respectively the discrete Fourier transforms of the reference frame and of the current frame.

5.3.2 Local Motion Compensation

A global motion model does not reflect the local interframe motions. Block-matching [135] is a standard procedure used in video processing to compensate for these local interframe motions. Its principle is illustrated in Figure 5.2. Here, each of the $\tau-1$ neighboring frames is replaced by a version that is motion-compensated w.r.t. the reference frame. Considering one of these neighboring frames, motion compensation is performed as follows: the reference frame is divided into blocks[2]; then, for each block of this frame, a search for similar blocks is performed in the neighboring frame; the compensated frame is then built by pasting the best matching block of the neighboring frame at the location of the reference block. Several parameters are therefore involved:

[2]In this work, we only consider non-overlapping blocks. Note that better PSNR results ($0.2-0.7$ dB) can be obtained with overlapping blocks, but the computational burden then becomes heavier.

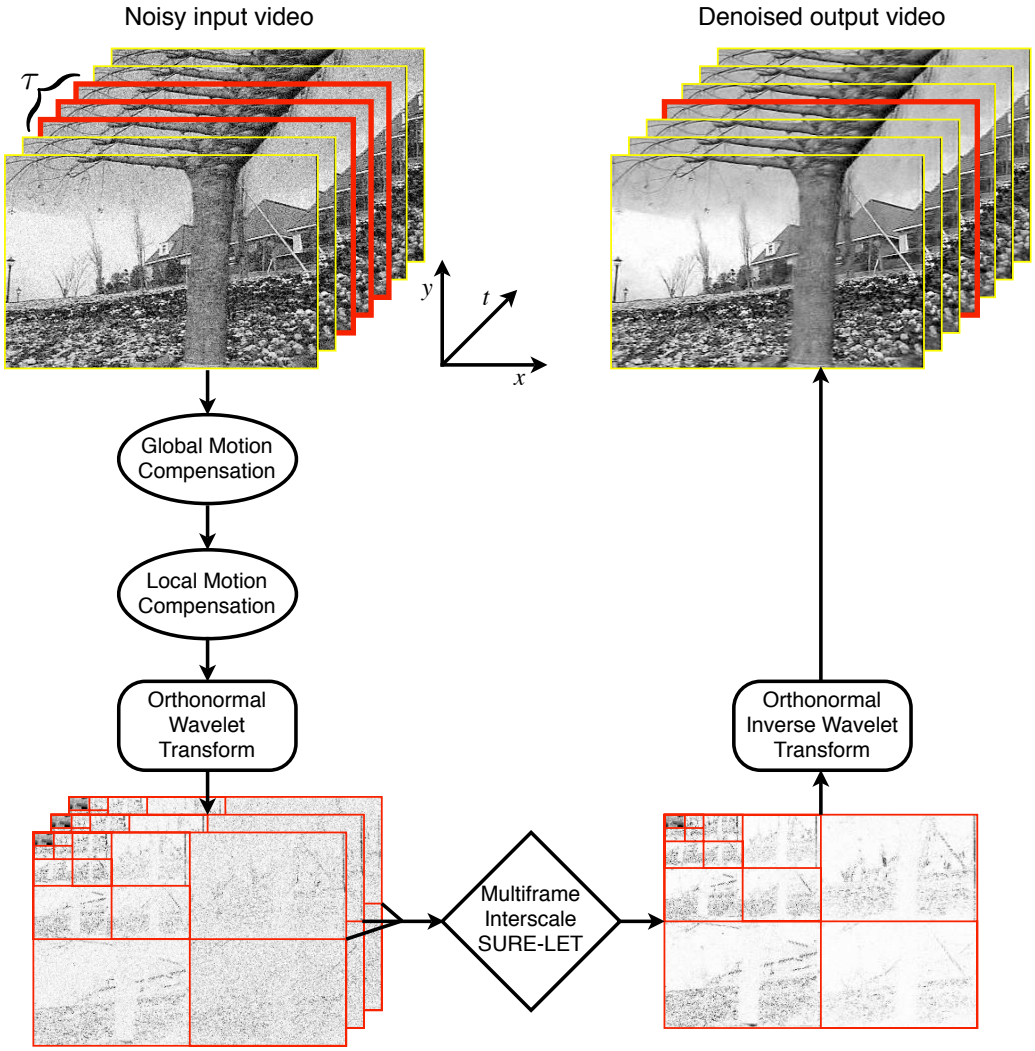


Figure 5.1: Overview of the proposed denoising algorithm. One frame of the video sequence is denoised by using its $\tau - 1$ neighbours (l.h.s., in red); here, $\tau = 3$. After compensating for their relative motion, these τ frames undergo an orthonormal wavelet transform, the coefficients of which are SURE-LET thresholded to provide a denoised estimate of the wavelet coefficients of the central frame.

- the size of the considered blocks: we found that rectangular blocks of fixed size 8×16 were a good trade-off between accurate motion estimation, robustness w.r.t. noise and computational complexity. Note that a rectangular shape is well-adapted to the standard video format, which are not of squared size.

- the size of the search region: here again, the trade-off evoked above led us to consider a square region of 15×15 pixels centered around the position of

the reference block. Note that we obtained similar results with a rectangular search region of 11×21 pixels.

- the criterion used for measuring the similarity between blocks: the two most popular measures of similarities are the mean of the absolute difference (MAD) and the mean squared error (MSE). We experimentally observed that the MSE gave slightly better results.

- the way of exploring the search region: we retained the exhaustive search because of its simplicity and accuracy. Note that there is a huge amount of literature (e.g. [137–139]) exposing fast algorithms for efficiently exploring the search region.

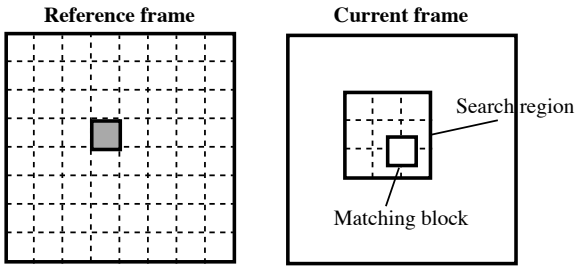


Figure 5.2: The principle of block-matching.

Instead of trying to find the best matches for every blocks of the reference frame, we consider only blocks where a significant motion occurred. Indeed, in noisy video sequences, there is a strong risk of matching the noise component in the still regions. In that case, the interframe noise becomes locally highly correlated (see Fig. 5.5(B)). To avoid this risk and still be able to consider the interframe noise as stationary (with a good approximation), we propose to perform motion compensation only in the blocks where a significant motion between frames was detected, as illustrated in Figure 5.3. Incidentally, this also significantly reduces the computation cost.

The proposed motion detection involves two steps:

1. In order to be robust w.r.t. noise, the considered frames are smoothed by the following regularized Wiener filter:

$$H(\omega) \quad = \quad \begin{cases} 1 - \frac{|N(\omega)|^2}{|U(\omega)|^2} & \text{, if } |U(\omega)|^2 > \lambda_1 |N(\omega)|^2 \\ 0 & \text{, otherwise} \end{cases} \tag{5.8}$$

where $|N(\omega)|^2$ and $|U(\omega)|^2$ are respectively the power spectrum of the noise (constant for white Gaussian noise) and of the noisy frame. $\lambda_1 \geq 1$ is the regularization parameter; its value will be discussed hereafter.

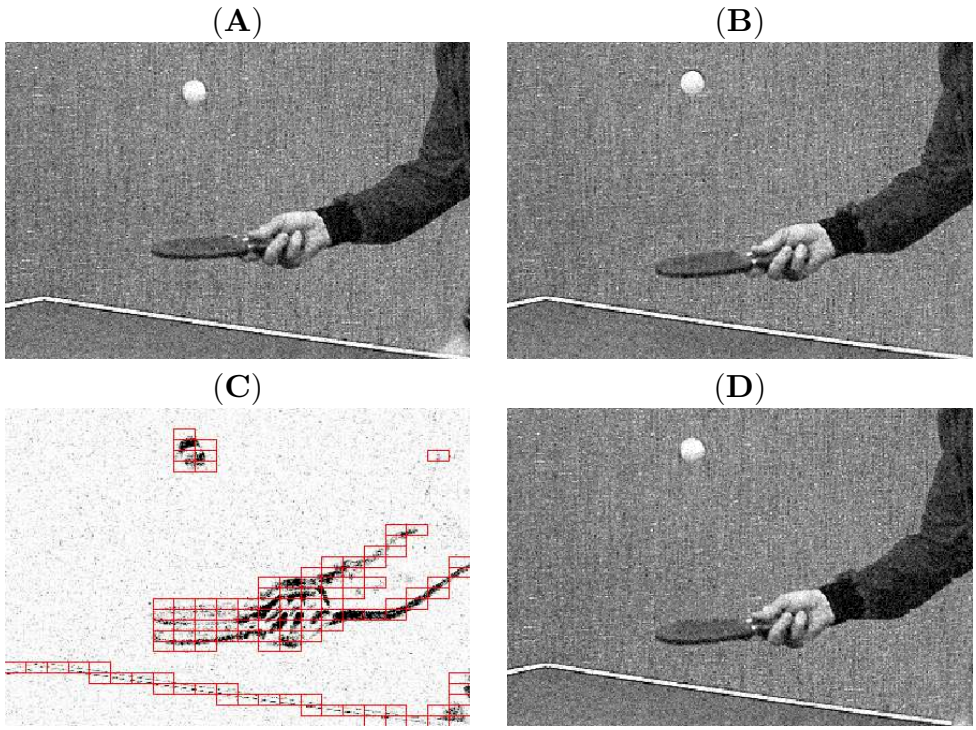


Figure 5.3: (**A**) Frame no 3 of the *Tennis* sequence: PSNR = 22.11 dB. (**B**) Frame no 6 of the *Tennis* sequence (reference frame): PSNR = 22.11 dB. (**C**) Detected motion with corresponding blocks to be matched. (**D**) Motion compensated frame no 3.

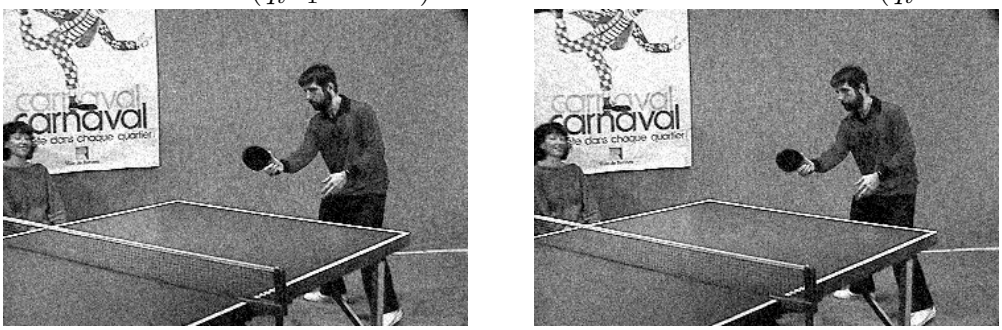
2. The MSEs between the two considered frames are then computed inside each block. The minimum of these MSEs (MSE_{min}) is considered as the "no motion level". Consequently, the block-matching will be performed only for those blocks of the reference frame having a MSE above a given threshold of motion $\lambda_2\ \text{MSE}_{\text{min}}$, where $\lambda_2 \geq 1$.

In our experiments, we found that any values of λ_1 and λ_2 chosen in the range $[2; 3]$ gave similar results (± 0.1 dB). A smaller value of these two parameters will decrease the robustness w.r.t. noise. A higher value of the regularization parameter λ_1 will oversmooth the frames, decreasing the accuracy of the subsequent block-matching. A higher value of the parameter λ_2 will speed up the algorithm, but the subsequent motion compensation will be less effective. In our experiments, we have selected $\lambda_1 = \lambda_2 = \sqrt{6}$.

The block-matching itself is performed on the smoothed frames, in order to decrease the sensitivity to noise. For each frame and for each detected block, the minimum MSE (computed between the reference block and its best matching block) is stored; the inverse of the average of these MSEs will then serve as a weight q_t for the

considered frame t in the subsequent wavelet-domain thresholding (Section 5.3.3). These weights are especially important when there is no or little correlation between adjacent frames; this situation appears when, for example, a quick change of camera occurs (see Figure 5.4).

Frame no 88 ($q_{t-1} = 0.69$) **Reference frame no 89** ($q_t = 0.69$)



Frame no 90 ($q_{t+1} = 0.21$)

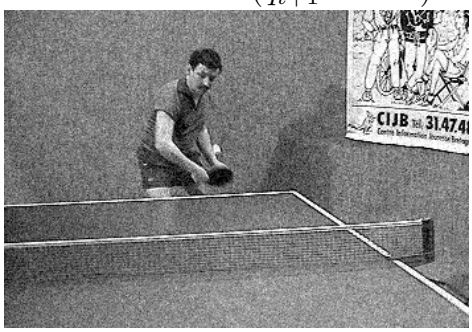


Figure 5.4: Example of a quick change of camera (*Tennis* sequence at PSNR = 22.11 dB). The various frames are weighted accordingly (q_{t-1}, q_t and q_{t+1}).

The proposed *selective* block-matching procedure has two key advantages:

1. It leads to a fast local motion compensation, despite the fact that an exhaustive search is performed.

2. The interframe noise covariance matrix can be assumed to be unaffected by the local motion compensation (Figure 5.5(C)), contrary to standard block-matching (Figure 5.5(B)).

5.3.3 Multiframe Interscale Wavelet Thresholding

Once the motion between a reference frame and a reasonable number of adjacent frames has been compensated, a 2D *orthonormal* wavelet transform is applied to each

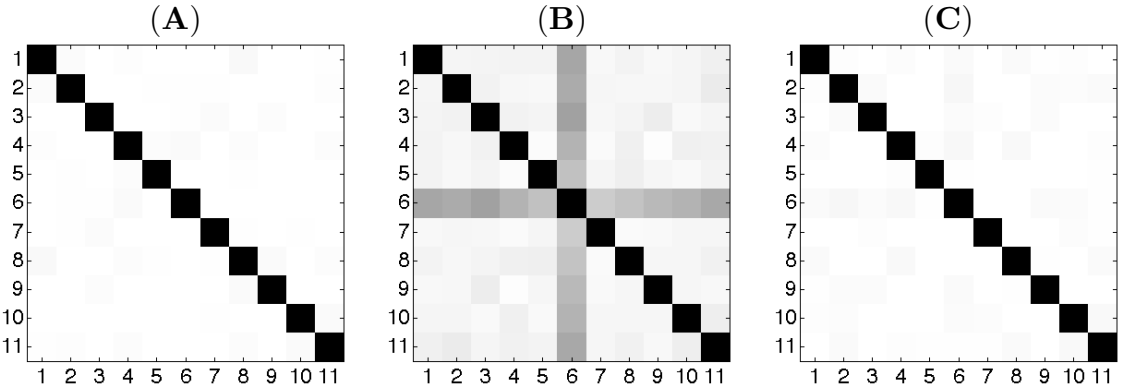


Figure 5.5: Influence of the block-matching motion compensation on the interframe noise covariance matrix (frame no 6 is the reference frame). (**A**) Interframe noise covariance matrix for the 11 first frames of the noisy *Tennis* sequence before motion compensation. (**B**) Interframe noise covariance matrix after a standard block-matching algorithm. (**C**) Interframe noise covariance matrix after the proposed selective block-matching algorithm.

motion-compensated frame. Each highpass subband of the reference frame is then denoised according to the generic procedure described in Section 5.2.2, Equ. (5.5), in which $K = 4$ and

$$\begin{aligned}\theta(\mathbf{y}_n, \tilde{\mathbf{y}}_n) \;=\; & \mathbf{a}_1^{\mathrm{T}} \underbrace{\gamma(\tilde{\mathbf{y}}_n^{\mathrm{T}}\mathbf{W}\tilde{\mathbf{y}}_n)\gamma(\mathbf{y}_n^{\mathrm{T}}\mathbf{W}\mathbf{y}_n)\mathbf{y}_n}_{\boldsymbol{\theta}_1(\mathbf{y}_n,\tilde{\mathbf{y}}_n)} + \\ & \mathbf{a}_2^{\mathrm{T}} \underbrace{\overline{\gamma}(\tilde{\mathbf{y}}_n^{\mathrm{T}}\mathbf{W}\tilde{\mathbf{y}}_n)\gamma(\mathbf{y}_n^{\mathrm{T}}\mathbf{W}\mathbf{y}_n)\mathbf{y}_n}_{\boldsymbol{\theta}_2(\mathbf{y}_n,\tilde{\mathbf{y}}_n)} + \\ & \mathbf{a}_3^{\mathrm{T}} \underbrace{\gamma(\tilde{\mathbf{y}}_n^{\mathrm{T}}\mathbf{W}\tilde{\mathbf{y}}_n)\overline{\gamma}(\mathbf{y}_n^{\mathrm{T}}\mathbf{W}\mathbf{y}_n)\mathbf{y}_n}_{\boldsymbol{\theta}_3(\mathbf{y}_n,\tilde{\mathbf{y}}_n)} + \\ & \mathbf{a}_4^{\mathrm{T}} \underbrace{\overline{\gamma}(\tilde{\mathbf{y}}_n^{\mathrm{T}}\mathbf{W}\tilde{\mathbf{y}}_n)\overline{\gamma}(\mathbf{y}_n^{\mathrm{T}}\mathbf{W}\mathbf{y}_n)\mathbf{y}_n}_{\boldsymbol{\theta}_4(\mathbf{y}_n,\tilde{\mathbf{y}}_n)} \end{aligned} \tag{5.9}$$

where

- $\gamma(x) = \exp\left(-\frac{|x|}{2\lambda_3^2}\right)$ and $\overline{\gamma}(x) = 1 - \gamma(x)$ are two discriminative functions that classify the wavelet coefficients in four groups, based on their magnitude and the magnitude of their parent $\tilde{\mathbf{y}}_n$. λ_3 is a threshold that rules this categorization of the wavelet coefficients. The numerous[3] linear parameters involved in the multiframe thresholding bring a high level of flexibility to the denoising

[3]Four times the considered number of adjacent frames per subband.

process. As a consequence, the nonlinear parameter λ_3 does not require a data-dependent optimization; we experimentally found that $\lambda_3 = \lambda_2 = \lambda_1 = \sqrt{6}$ gave the best results. Note that this value is the same that we used for multichannel denoising in Section 3.4.3 of Chapter 3.

- Each $\mathbf{a}_k$ is a $\tau \times 1$ vector of linear parameters that is optimized for each subband by the procedure described in Section 5.2.2.

- $\mathbf{W} = \mathbf{Q}^{\mathrm{T}} \mathbf{\Gamma}_\tau^{-1} \mathbf{Q}$ is a $\tau \times \tau$ weighting matrix that takes into account:

  1. the potential interframe SNR disparities, through the inverse of the $\tau \times \tau$ interframe noise covariance matrix $\mathbf{\Gamma}_\tau$;
  2. the potential weak interframe correlations, through the weights q_t resulting from the block-matching (Section 5.3.2) and stored in the $\tau \times \tau$ diagonal matrix $\mathbf{Q}$. The weights are normalized to ensure that the Frobenius norm of $\mathbf{Q}$, $\|\mathbf{Q}\|_F = \sqrt{\mathrm{trace}\,\{\mathbf{Q}^T \mathbf{Q}\}} = 1$.

5.3.4 Computational Complexity

To denoise a given reference frame using its $(\tau - 1)$ neighboring frames, the computational complexity of our algorithm can be evaluated as follows:

- Discrete wavelet transform: $\mathcal{O}(\tau \cdot N \cdot \log_2(N))$
- Building of the interscale predictor: $\mathcal{O}(\tau \cdot N \cdot \log_2(N))$
- Application of the interscale wavelet thresholding (5.9): $\mathcal{O}(\tau \cdot K \cdot N)$
- Global motion compensation: $\mathcal{O}((\tau - 1) \cdot N \cdot \log_2(N))$
- Local motion compensation: $\mathcal{O}((\tau - 1) \cdot B_x \cdot B_y \cdot R_x \cdot R_y \cdot \frac{N}{S_x \cdot S_y})$, where $B_x \times B_y$ is the block size, $R_x \times R_y$ the size of the search region and (S_x, S_y) the step size between two adjacent reference blocks.

In the proposed settings, we consider $\tau = 11$ adjacent frames (5 past , 5 future and the current frame), non-overlapping —$(S_x, S_y) = (B_x, B_y)$— blocks of size $(B_x, B_y) = (8, 16)$, a search region of $(R_x, R_y) = (15, 15)$ locations and $K = 4$ vectors of $\tau \times 1$ parameters. For a frame of $N = 288 \times 352$ pixels, the most costly part of our algorithm is thus the local motion compensation step, because $(\tau - 1) R_x R_y >> \tau \log_2(N) >> \tau K$. When summing up all these operations, we get an approximate number of 2800 operations per pixel to denoise one frame. However, since the block matching procedure is selective, the actual number of operations is much lower in practice.

5.4 Experiments

We propose now to evaluate the performance of our algorithm in comparison to some other state-of-the-art *wavelet-based* methods (all are redundant):

- Selesnick *et al.* 3DWTF [84, 127]: a bivariate thresholding is applied in each spatio-temporal subband of a 3D non-separable dual-tree complex wavelet transform.
- Pižurica *et al.* SEQWT [128]: a spatially adaptive Bayesian shrinkage is applied in the undecimated wavelet-domain, followed by a recursive temporal filtering.
- Zlokolica *et al.* WRSTF [129]: in the undecimated wavelet domain, motion estimation and adaptive temporal filtering are recursively performed, followed by an intraframe spatially adaptive filter.

The results of these three algorithms for denoising standard grayscale video sequences can be downloaded at `http://telin.ugent.be/~vzlokoli/Results_J/`. This allows a fair comparison between the various methods. Note that we could not directly compare our results with the recent multiframe extension [133] of the original *BLS-GSM* algorithm [81], because we could neither have access to the corresponding code nor strictly reproduce their experiments. However, we observed a similar improvement over the still *BLS-GSM* with our method and with the algorithm described in [133].

The noisy video sequences have been obtained by adding (without clipping) independent Gaussian white noises of given variance σ^2 on each frame, i.e. $\mathbf{\Gamma} = \sigma^2 \mathbf{Id}$. For our algorithm, we performed 4 levels of decomposition of an *orthonormal* wavelet transform using Daubechies *sym8* filters [37].

In Figure 5.6, we show the peak signal-to-noise ratio[4] (PSNR) in each frame of various video sequences[5] (*Flowers*, *Tennis*, *Foreman* and *Bus*) at different input PSNR. We can observe that our algorithm achieves globally, and for almost every frames, significantly better results than the three other state-of-the-art procedures. From a visual point of view, our solution provides a good trade-off between noise reduction and preservation of small features (see Figure 5.7), as confirmed by the SSIM index.

In Table 5.1, we show global PSNR comparisons of the various wavelet-based algorithms. As observed, the proposed algorithm consistently gives higher PSNR (+1 dB) than the other wavelet-based approaches. We must point out that these denoising results are very encouraging since we are only using a non-redundant (orthonormal) wavelet transform, whereas other state-of-the-art video denoising algorithms make use of *redundant* wavelet representations. Moreover, with our current

[4] Defined as: $\mathrm{PSNR} = 10 \log_{10} \frac{255^2}{\mathrm{MSE}}$ dB.

[5] Available together with some denoising results at:
`http://bigwww.epfl.ch/luisier/VideoDenoising/`

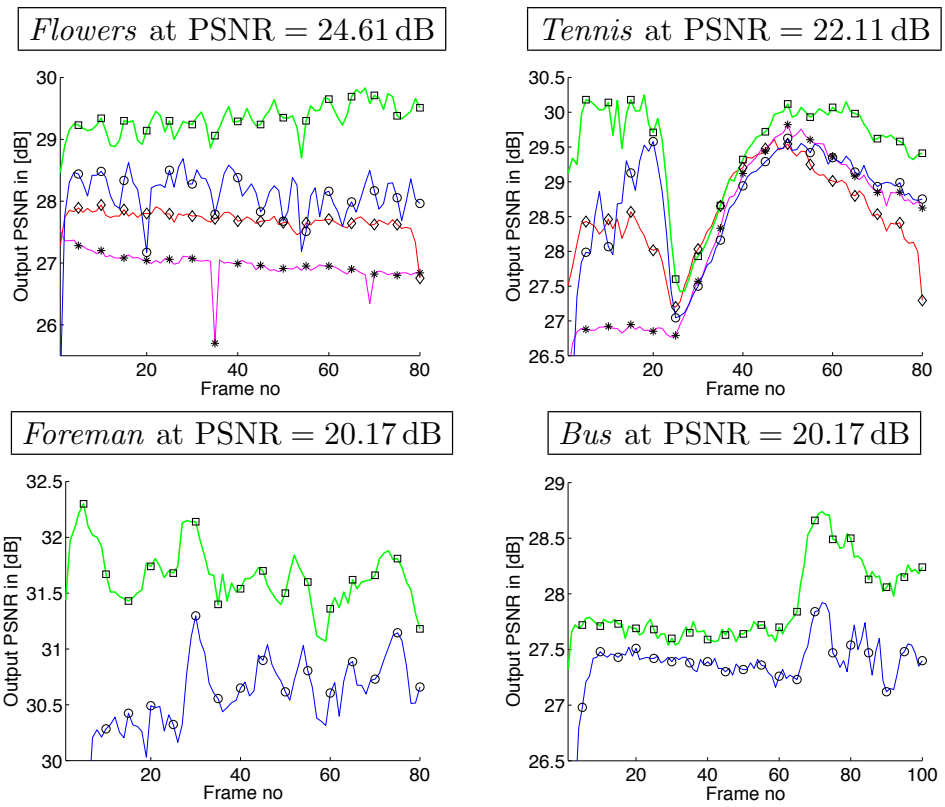


Figure 5.6: Comparison of the PSNR evolution for various video sequences and wavelet-based denoising algorithms. "□" markers refer to the proposed algorithm, "◇" to the *3DWTF* [83, 127], "∗" to the *SEQWT* [128] and "○" to the *WRSTF* [129].

Matlab implementation, it only takes around 2s to denoise one 288×352 frame. This is slightly slower than the execution time of the *WRSTF*, but the latter is fully implemented in C++.

We have also added in Table 5.1 the results we obtained by running the *VBM3D*[6] [140], which is not a wavelet-based solution, but a more sophisticated two-pass algorithm. Up to our knowledge, the PSNRs obtained by the *VBM3D* are among the best published so far for video denoising [141]. Note that the results presented in [141] could not be reproduced due to unspecified video clipping and cropping[7]. We can notice in Table 5.1 that, in some cases, our *lower-complexity* solution is even competitive with the *VBM3D*.

From a computational standpoint, our algorithm only requires $\sim 2'800$ operations per pixel (see Section 5.3.4). To compare with, the algorithm described in [141]

[6]Matlab code available at: `http://www.cs.tut.fi/~foi/GCF-BM3D/#ref_software`
[7]Private communication with the authors of [141].

Table 5.1: Comparison of some state-of-the-art video denoising algorithms

σ	5	10	15	20	25	30	50	100
Input PSNR	34.15	28.13	24.61	22.11	20.17	18.59	14.15	8.13
Method	**Tennis 240 × 352**							
3DWTF [83, 127]	N/A	32.09	29.88	28.31	N/A	N/A	N/A	N/A
SEQWT [128]	N/A	30.96	28.34	26.88	N/A	N/A	N/A	N/A
WRSTF [129]	N/A	32.57	30.45	28.96	27.83	N/A	N/A	N/A
Proposed Algorithm	37.38	**33.52**	**31.41**	**29.96**	**28.85**	27.94	25.50	22.60
VBM3D [121]	*37.82*	*34.14*	*32.10*	*30.39*	*28.84*	*27.41*	*24.81*	*22.13*
Method	**Flowers 240 × 352**							
3DWTF [83, 127]	N/A	30.33	27.79	26.10	N/A	N/A	N/A	N/A
SEQWT [128]	N/A	29.62	27.10	25.32	N/A	N/A	N/A	N/A
WRSTF [129]	N/A	30.77	28.10	26.33	24.92	N/A	N/A	N/A
Proposed Algorithm	36.22	**31.63**	**29.11**	**27.31**	**25.89**	24.76	21.73	18.54
VBM3D [121]	*36.51*	*32.04*	*29.66*	*28.07*	*26.82*	*25.74*	*21.53*	*17.27*
Method	**Foreman 288 × 352**							
WRSTF [129]	N/A	35.33	33.14	31.55	30.30	N/A	N/A	N/A
Proposed Algorithm	39.60	**36.13**	**34.13**	**32.73**	**31.61**	30.71	28.15	24.85
VBM3D [121]	*40.21*	*37.19*	*35.59*	*34.39*	*33.39*	*32.52*	*29.75*	*24.02*
Method	**Bus 288 × 352**							
WRSTF [129]	N/A	32.78	30.40	28.76	27.46	N/A	N/A	N/A
Proposed Algorithm	37.45	**33.13**	**30.69**	**28.99**	**27.74**	26.74	24.04	21.12
VBM3D [121]	*37.28*	*32.88*	*30.49*	*28.95*	*27.81*	*26.93*	*24.40*	*20.74*

Note: PSNRs displayed in this table correspond to the averaged values over frames $10-20$ of the various video sequences, using frames $5-25$ to avoid potential boundary artifacts in the temporal dimension.

requires $\sim 75'000$ operations per pixel. Note that the authors in [141] have suggested some acceleration techniques which could potentially reduce this number of operations per pixel. In [140], there is no analysis of algorithm's complexity. However, in their paper on image denoising [121], they give the following formula for computing the overall number of operations per pixel (using our notations):

$$\text{Cost} = 3C_{\text{2D}} + 2\frac{(B_xB_y + N_2)R_xR_y}{S_xS_y} + 3\frac{(N_2C_{\text{2D}} + B_xB_yC_{\text{1D}})}{S_xS_y}$$

where N_2 is the number of blocks in a 3D group, C_{1D} and C_{2D} denote the number of arithmetic operations required for the 1D (resp. 2D) transform of the group of blocks.

In their video denoising settings, they use $(B_x, B_y) = (8, 8)$, $(R_x, R_y) = (7, 7)$, $(S_x, S_y) = (6, 6)$, $N_2 = 8$ and $\tau = 9$ neighboring frames. Considering a separable FFT-based implementation, the cost of the 2D (resp. 1D) transform can be evaluated as $C_{\text{2D}} = B_xB_y \log_2(B_xB_y)$ (resp. $C_{\text{1D}} = N_2 \log_2(N_2)$). Using these values, we obtain an overall number of $\tau \times \text{Cost} \simeq 15'500$ operations per pixel to denoise one frame. Note that they have also proposed some acceleration techniques in order to decrease this number of operations.

5.5 Summary

In this chapter, we have presented a relatively simple and yet very efficient *orthonormal* wavelet-domain video denoising algorithm. Thanks to a proper selective block-matching procedure, the effect of motion compensation on the noise statistics becomes negligible, and an adapted multiframe interscale SURE-LET thresholding could be applied. The proposed algorithm has been shown to favorably compare with state-of-the-art *redundant* wavelet-based approaches, while having a lighter computational load. The consideration of redundant transformations seems however necessary for our solution to consistently reach the same level of performance as the very best video denoising algorithms available [121, 141].

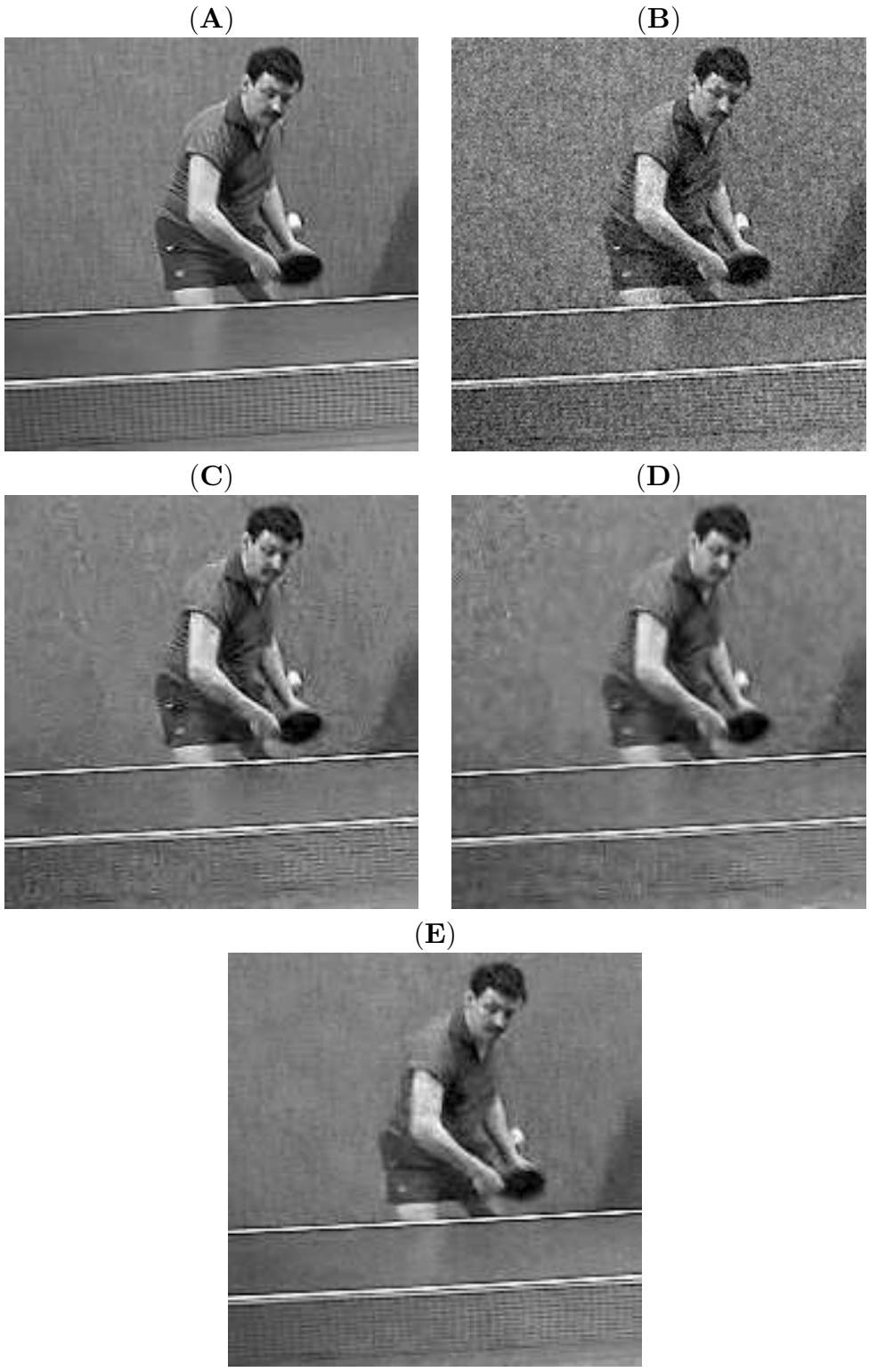


Figure 5.7: (**A**) Part of the noise-free frame no 90 of the *Tennis* sequence. (**B**) Noisy version of it: SSIM = 0.46 (**C**) Denoised with the *SEQWT* [128]: SSIM = 0.71. (**D**) Denoised with the *WRSTF* [129]: SSIM = 0.69. (**E**) Denoised with the proposed algorithm: SSIM = 0.82.

Part II

Poisson Intensities Estimation in AWGN

Chapter 6

PURE-LET: a Toolkit for Estimating Poisson Intensities in AWGN

6.1 Motivations

"Additive white Gaussian noise" (AWGN) is a ubiquitous model in the context of statistical image restoration. In many applications, however, the current trend towards quantitative imaging calls for less generic models that better account for the physical acquisition process. The need for such models is particularly stringent in biomicroscopy, where live samples are often observed at very low light levels, due to acquisition-time and photo-toxicity constraints [142]. In this regime, the performance of the imaging device is typically shot-noise limited, i.e. the major source of measurement noise is strongly signal-dependent. Thus, opting for a non-additive, non-Gaussian model can yield significant restoration-quality improvements in such applications.

Gaussian priors almost systematically lead to closed-form solutions in a Bayesian framework; this has probably led to the common misconception that other noise models are generally less tractable. Actually, when using non-Bayesian approaches, the advantage of the Gaussian paradigm is not as clear-cut. In this chapter, we provide a rigorous data-driven procedure, hereafter termed *PURE-LET*, for estimating Poisson intensities in AWGN. The driving principle of the proposed approach still remains the minimization of the MSE.

6.2 Related Work

The first wavelet-based techniques that were specifically designed for Poisson intensity estimation appeared in the fields of astrophysics [143] and seismology [144]. In the context of image denoising, several works are based on the fact that Poisson statistics are preserved across scales in the lowpass channel of an unnormalized Haar wavelet transform. This fundamental property was for example used by Timmermann and Nowak [145] to construct a multiscale Bayesian model of the image; an extension for estimating all parameters of the model using the expectation-maximization algorithm was derived in [146]. A similar model was proposed independently by Kolaczyk for 1D signals [147], using the concept of recursive dyadic partition. A generalization to three families of distributions (Gaussian, Poisson and multinomial) was described in [148], along with an asymptotic minimax analysis. The aforementioned property was also used within a user-calibrated hypothesis-testing approach for astrophysical imaging [149]. The idea of using hypothesis testing to handle Poisson statistics was initially adapted from the Gaussian case by Kolaczyk [150], who proposed to use (pairs of) level-dependent thresholds. A complementary study of the theoretical asymptotic properties of wavelet estimators was presented in [151].

Notice that the Bayesian framework has also been used in conjunction with more involved multiscale transformations than the Haar transform. For example, Sardy et al. [152] considered a general ℓ_1-penalized-likelihood framework for arbitrary wavelet bases and noise models, including the Poisson case. More recently, Willett and Nowak have proposed a platelet-based penalized-likelihood estimator that was shown to be very efficient for denoising piecewise-smooth images [153].

A widespread alternative to the direct handling of Poisson statistics is to apply variance-stabilizing transforms (VSTs), with the underlying idea of exploiting the broad class of denoising methods that are based on a Gaussian noise model [154]. Since the seminal work of Anscombe [155], more involved VSTs have been proposed, the most famous example being the Haar-Fisz transform [156]. Such approaches belong to the state-of-the-art for 1D wavelet-based Poisson noise removal [156, 157]. They have been combined with various other methodologies, e.g. Bayesian multiscale likelihood models that can be applied to arbitrary wavelet transforms [158]. Very recently, a hybrid approach that combines VSTs, hypothesis testing, ℓ_1-penalized reconstruction and advanced redundant multiscale representations has been proposed by Zhang et al. [159]. Their approach and the *Platelet* method of Willett and Nowak, stand among the most efficient estimators of 2D Poisson intensities.

With the exception of cross-validation methods [152, 160, 161], however, the potential of purely data-driven techniques seems to have remained under-exploited for the wavelet-based restoration of images corrupted by Poisson noise.

6.3 Poisson Unbiased Risk Estimate (PURE)

Recall that:

- z is a Poisson random variable of intensity $x \in \mathbb{R}_+$, if and only if its probability density function (PDF) p conditioned on x is

$$p(z|x) = \frac{x^z}{z!} e^{-x}$$

  We use the standard notation $z \sim \mathcal{P}(x)$.

- b is a normally distributed random variable with zero-mean and variance σ^2, if and only if its probability density function q is

$$q(b) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{b^2}{2\sigma^2}}$$

  We use the standard notation $b \sim \mathcal{N}(0, \sigma^2)$.

In this chapter, $\mathbf{z}$ will denote a vector of N independent Poisson random variables z_n of underlying intensities x_n, i.e. $z_n \sim \mathcal{P}(x_n)$. A realization of $\mathbf{z}$ can be thought of as a noisy measurement of the intensity signal $\mathbf{x}$. We recall that, in contrast with Bayesian approaches, $\mathbf{x}$ is considered to be *deterministic* in the present framework.

Now, let us further assume that these noisy measurements z_n are degraded by i.i.d. additive white Gaussian noises (AWGNs) b_n of known variance σ^2, i.e. $b_n \sim \mathcal{N}(0, \sigma^2)$. The final observation vector $\mathbf{y}$ is therefore given by:

$$\mathbf{y} = \mathbf{z} + \mathbf{b} \tag{6.1}$$

Our goal is then to find an estimate $\hat{\mathbf{x}} = \mathbf{F}(\mathbf{y})$ that is the closest possible to the unknown original signal in the minimum mean-squared error (as defined in Equ. (2.2)) sense. In order to build a purely *data-driven* unbiased estimate of the MSE, we need to consider the following lemma:

Lemma 4. *Let the family of vectors $(\mathbf{e}_n)_{n=1\ldots N}$ be the canonical basis of $\mathbb{R}^N$, i.e. all components of $\mathbf{e}_n$ are zero, except for the n^{th} component, which is equal to one. Let $\mathbf{F} : \mathbb{R}^N \to \mathbb{R}^N$ be a real vector-valued function such that $\forall n$, $\mathscr{E}_{\mathbf{z}} \{|f_n(\mathbf{z} - \mathbf{e}_n)|\} < \infty$ and $f_n(\mathbf{z}) = 0$ if $z_n \leq 0$. Then*

$$\mathscr{E}_{\mathbf{z}} \left\{ \mathbf{x}^{\mathrm{T}} \mathbf{F}(\mathbf{z}) \right\} = \mathscr{E}_{\mathbf{z}} \left\{ \mathbf{z}^{\mathrm{T}} \mathbf{F}^{-}(\mathbf{z}) \right\} \tag{6.2}$$

where $\mathbf{F}^{-}(\mathbf{z}) = [f_n(\mathbf{z} - \mathbf{e}_n)]_{n=1\ldots N}$.

Proof. This result can be thought of as the "Poisson equivalent" of Stein's lemma. A proof of a similar result can be found for instance in [162, 163]. Yet, for the sake of completeness, we give this proof hereafter.

Denoting by $\mathscr{E}_{z_n}\{\cdot\}$ the partial expectation over the n^{th} component of the Poisson process $\mathbf{z}$, we have the following sequence of equalities:

$$\begin{aligned}
\mathscr{E}_{z_n}\left\{x_n f_n(\mathbf{z})\right\} &= \sum_{z_n=0}^{+\infty} x_n f_n(\mathbf{z}) p(z_n|x_n) \\
&= \sum_{z_n=0}^{+\infty} x_n f_n(\mathbf{z}) \frac{x_n^{z_n}}{z_n!} e^{-x_n} \\
&= \sum_{z_n=1}^{+\infty} f_n(\mathbf{z}-\mathbf{e}_n) \frac{x_n^{z_n}}{(z_n-1)!} e^{-x_n} \\
&= \sum_{z_n=0}^{+\infty} z_n f_n(\mathbf{z}-\mathbf{e}_n) \frac{x_n^{z_n}}{z_n!} e^{-x_n} \\
&= \mathscr{E}_{z_n}\left\{z_n f_n(\mathbf{z}-\mathbf{e}_n)\right\}
\end{aligned}$$

Taking the expectation over the remaining components of $\mathbf{z}$ and using the linearity of the mathematical expectation finally leads to the desired result. ■

The MSE score of a given estimate $\hat{\mathbf{x}} = \mathbf{F}(\mathbf{y})$ of $\mathbf{x}$ can then be accurately evaluated from the observed data using the following theorem:

Theorem 4. *Let $\mathbf{F}(\mathbf{y}) = [f_n(\mathbf{y})]_{n=1\ldots N}$ be a N-dimensional real-valued vector function satisfying the requirements of Lemma 4. Then the random variable*

$$\begin{aligned}
\epsilon &= \frac{1}{N}\left(\|\mathbf{F}(\mathbf{y})\|^2 - 2\mathbf{y}^{\mathrm{T}}\mathbf{F}^{-}(\mathbf{y}) + 2\sigma^2 \mathrm{div}\left\{\mathbf{F}^{-}(\mathbf{y})\right\}\right) + \\
&\quad \frac{1}{N}\left(\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right) - \sigma^2
\end{aligned} \tag{6.3}$$

is an unbiased *estimate of the expected MSE, i.e.*

$$\mathscr{E}\{\epsilon\} = \frac{1}{N}\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y}) - \mathbf{x}\|^2\right\}$$

Proof. By expanding the expectation of the squared error between $\mathbf{x}$ and its estimate $\mathbf{F}(\mathbf{y})$, we obtain

$$\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y}) - \mathbf{x}\|^2\right\} = \mathscr{E}\left\{\|\mathbf{F}(\mathbf{y})\|^2\right\} + \underbrace{\|\mathbf{x}\|^2}_{\text{(I)}} - 2\underbrace{\mathscr{E}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\right\}}_{\text{(II)}} \tag{6.4}$$

We can now evaluate the two expressions (I,II) which involve the unknown data $\mathbf{x}$.

(I) We notice that

$$\begin{aligned}\|\mathbf{x}\|^2 &= \mathscr{E}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{y}\right\}\\ &\overset{(6.1)}{=} \mathscr{E}_{\mathbf{b}}\left\{\mathscr{E}_{\mathbf{z}}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{z}\right\}\right\} + \underbrace{\mathscr{E}_{\mathbf{b}}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{b}\right\}}_{=0}\\ &\overset{\text{Lemma 4}}{=} \mathscr{E}_{\mathbf{b}}\left\{\mathscr{E}_{\mathbf{z}}\left\{\|\mathbf{z}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{z}\right\}\right\}\end{aligned}$$

and we have:

$$\begin{aligned}\mathscr{E}_{\mathbf{b}}\left\{\|\mathbf{z}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{z}\right\} &\overset{(6.1)}{=} \mathscr{E}_{\mathbf{b}}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} + \mathscr{E}_{\mathbf{b}}\left\{\|\mathbf{b}\|^2 - 2\mathbf{b}^{\mathrm{T}}\mathbf{y}\right\}\\ &\overset{\text{Lemma 1}}{=} \mathscr{E}_{\mathbf{b}}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} - N\sigma^2\end{aligned}$$

which finally leads to:

$$\|\mathbf{x}\|^2 = \mathscr{E}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} - N\sigma^2 \tag{6.5}$$

(II) We can write the following sequence of equalities:

$$\begin{aligned}\mathscr{E}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\right\} &= \mathscr{E}_{\mathbf{b}}\left\{\mathscr{E}_{\mathbf{z}}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{F}(\mathbf{z}+\mathbf{b})\right\}\right\}\\ &\overset{\text{Lemma 4}}{=} \mathscr{E}_{\mathbf{b}}\left\{\mathscr{E}_{\mathbf{z}}\left\{\mathbf{z}^{\mathrm{T}}\mathbf{F}^{-}(\mathbf{z}+\mathbf{b})\right\}\right\}\\ &\overset{(6.1)}{=} \mathscr{E}\left\{\mathbf{y}^{\mathrm{T}}\mathbf{F}^{-}(\mathbf{y})\right\} - \mathscr{E}_{\mathbf{z}}\left\{\mathscr{E}_{\mathbf{b}}\left\{\mathbf{b}^{\mathrm{T}}\mathbf{F}^{-}(\mathbf{y})\right\}\right\}\\ &\overset{\text{Lemma 1}}{=} \mathscr{E}\left\{\mathbf{y}^{\mathrm{T}}\mathbf{F}^{-}(\mathbf{y})\right\} - \sigma^2\mathscr{E}\left\{\operatorname{div}\left\{\mathbf{F}^{-}(\mathbf{y})\right\}\right\}\end{aligned} \tag{6.6}$$

Putting back Equ. (6.5) and (6.6) into Equ. (6.4) finally demonstrates Theorem 4. ■

In the remaining of this work, the unbiased MSE estimate (6.3) will be referred to as PURE (Poisson's unbiased risk estimate)[1]. It can be seen as the equivalent of SURE (Stein's unbiased risk estimate) [70], which was derived for AWGN only.

In the following proposition, we give an extension to the Poisson's Lemma 4, that will be useful for our study on the reliability of PURE given in the next section.

Proposition 5. *Let $k \in \mathbb{N}^*$ be a positive integer and $f_n : \mathbb{R}^N \to \mathbb{R}$ real-valued functions such that, for $n = 1 \dots N$ and $i = 1 \dots k$, $|\mathscr{E}\{f_n(\mathbf{z} - i\mathbf{e}_n)\}| < +\infty$ and $f_n(\mathbf{z}) = 0$, if $z_n < k$. Then,*

$$\mathscr{E}\left\{x_n^k f_n(\mathbf{z})\right\} = \mathscr{E}\left\{\prod_{i=0}^{k-1}(z_n - i) f_n(\mathbf{z} - k\mathbf{e}_n)\right\} \tag{6.7}$$

The proof of this proposition can be derived in the same way as the proof of Lemma 4. Note that a similar result can be found in [163].

[1] This unbiased MSE estimate was first presented (without proof) in [164].

6.4 Reliability of the PURE Estimate

In this section, we propose to compute the expected squared error between the actual MSE and its PURE estimate. For the sake of simplicity, we only consider the case of a pure Poisson process and we do not take into account the error induced by the estimation of the noise-free signal energy $\|\mathbf{x}\|^2$.

Property 2. *In the case of a Poisson process, the expected squared error between the effective part of PURE* (6.3) *and the actual MSE is given by:*

$$\begin{aligned}
&\mathscr{E}\left\{(\epsilon - \mathrm{MSE})^2\right\} \\
&\quad = \frac{4}{N^2}\mathscr{E}\Bigg\{ \sum_{1\leq i,j\leq N} y_i y_j \Big(f_i(\mathbf{y}-\mathbf{e}_i) f_j(\mathbf{y}-\mathbf{e}_j) + f_i(\mathbf{y}-\mathbf{e}_i-\mathbf{e}_j) f_j(\mathbf{y}-\mathbf{e}_i-\mathbf{e}_j) \\
&\qquad -2 f_i(\mathbf{y}-\mathbf{e}_i) f_j(\mathbf{y}-\mathbf{e}_i-\mathbf{e}_j) \Big) \\
&\qquad + \sum_{1\leq i\leq N} y_i f_i(\mathbf{y}-2\mathbf{e}_i) \Big(2 f_i(\mathbf{y}-\mathbf{e}_i) - f_i(\mathbf{y}-2\mathbf{e}_i) \Big) \Bigg\}
\end{aligned} \tag{6.8}$$

Proof. The expected squared error can be expanded as:

$$\begin{aligned}
\frac{N^2}{4}\mathscr{E}\left\{(\epsilon - \mathrm{MSE})^2\right\} &= \underbrace{\mathscr{E}\left\{\left(\mathbf{x}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\right)^2\right\}}_{\text{(I)}} - 2\underbrace{\mathscr{E}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\mathbf{y}^{\mathrm{T}}\mathbf{F}^{-}(\mathbf{y})\right\}}_{\text{(II)}} \\
&\quad + \mathscr{E}\left\{\left(\mathbf{y}^{\mathrm{T}}\mathbf{F}^{-}(\mathbf{y})\right)^2\right\}
\end{aligned} \tag{6.9}$$

Only the expectations (I) and (II) involve the unknown Poisson intensities $\mathbf{x}$. Using Proposition 5, we can re-express them as:

$$\begin{aligned}
\text{(I)}\quad \mathscr{E}\left\{\left(\mathbf{x}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\right)^2\right\} &= \sum_{i,j} \mathscr{E}\left\{x_i x_j f_i(\mathbf{y}) f_j(\mathbf{y})\right\} \\
&= \sum_{i,j} \mathscr{E}\left\{y_i y_j f_i(\mathbf{y}-\mathbf{e}_i-\mathbf{e}_j) f_j(\mathbf{y}-\mathbf{e}_i-\mathbf{e}_j)\right\} - \\
&\quad \sum_{i} \mathscr{E}\left\{y_i f_i^2(\mathbf{y}-2\mathbf{e}_i)\right\}
\end{aligned}$$

$$\begin{aligned}
\text{(II)}\quad \mathscr{E}\left\{\mathbf{x}^{\mathrm{T}}\mathbf{F}(\mathbf{y})\mathbf{y}^{\mathrm{T}}\mathbf{F}^{-}(\mathbf{y})\right\} &= \sum_{i,j} \mathscr{E}\left\{x_i y_j f_i(\mathbf{y}) f_j(\mathbf{y}-\mathbf{e}_j)\right\} \\
&= \sum_{i,j} \mathscr{E}\left\{y_i y_j f_i(\mathbf{y}-\mathbf{e}_i) f_j(\mathbf{y}-\mathbf{e}_i-\mathbf{e}_j)\right\} - \\
&\quad \sum_{i} \mathscr{E}\left\{y_i f_i(\mathbf{y}-\mathbf{e}_i) f_i(\mathbf{y}-2\mathbf{e}_i)\right\}
\end{aligned}$$

The desired result (6.8) is finally demonstrated by putting back (I) and (II) into (6.9). ■

The expression (6.8) is usually very time-consuming to compute for an arbitrary nonlinear processing $\mathbf{F}$, because the computation of all components $f_i(\mathbf{y} - \mathbf{e}_i - \mathbf{e}_j)$, $i = 1 \dots N$, requires the order of N^2 applications of the whole denoising process $\mathbf{F}$. However, in the case of a *linear* processing $\mathbf{F}(\mathbf{y}) = \mathbf{A}\mathbf{y}$, where $\mathbf{A}$ is an arbitrary $N \times N$ matrix, the expected squared error between the effective part of PURE and the actual MSE becomes:

$$\begin{aligned} &\mathscr{E}\left\{(\epsilon - \mathrm{MSE})^2\right\} \\ &\quad = \frac{4}{N^2}\mathscr{E}\left\{\|\mathbf{F}(\mathbf{y})\|^2 + \mathbf{y}^{\mathrm{T}}\mathbf{F}(\mathbf{y})^2 - 2\mathbf{y}^{\mathrm{T}}\mathrm{diag}\left\{\mathbf{A}\right\}^{\mathrm{T}}\mathrm{diag}\left\{\mathbf{F}(\mathbf{y})\mathbf{y}^{\mathrm{T}}\right\}\right\} \end{aligned} \tag{6.10}$$

where $\mathbf{F}(\mathbf{y})^2 = [f_n^2(\mathbf{y})]_{1 \le n \le N}$.

If we further assume that $\mathbf{A}$ has the same structure as the one discussed in Section 2.4 of Chapter 2, i.e. $\mathbf{A}$ specifies a *pointwise* shrinkage in an orthonormal transform-domain, then all the terms which appear in (6.10) linearly grow with N. Consequently, the overall squared error decreases linearly as the number of measurements N increases. The reliability of PURE as an estimate of the MSE for a Poisson process is thus similar to the reliability of SURE discussed in Section 2.4.

6.5 Optimal/PURE-LET Estimator

As in the AWGN case, it is possible to find the processing $\mathbf{F}_{\mathrm{opt}}$ that minimizes the expected MSE/PURE, without an explicit knowledge of the PDF of the original noise-free signal $p(\mathbf{x})$. We denote the joint density of $\mathbf{x}$ and $\mathbf{y}$ by:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \tag{6.11}$$

and thus the marginal PDF of the noisy observation is $r(\mathbf{y}) = \displaystyle\int_{\mathbb{R}_+^N} p(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{x}$.

Theorem 5. *The optimal, in the minimum expected MSE/PURE sense, estimate of the intensity of a Poisson process embedded in an AWGN of variance σ^2 is given by:*

$$\mathbf{F}_{\mathrm{opt}}(\mathbf{y}) = \left[(y_n + 1)\frac{r(\mathbf{y} + \mathbf{e}_n)}{r(\mathbf{y})} + \frac{\sigma^2}{r(\mathbf{y})}\frac{\partial r(\mathbf{y} + \mathbf{e}_n)}{\partial y_n}\right]_{1 \le n \le N} \tag{6.12}$$

Proof. Taking the global expectation of PURE (6.3) leads to:

$$\begin{aligned} N\mathscr{E}\{\epsilon\} &= \int_{\mathbb{R}_+^N}\int_{\mathbb{R}^N} \left(\|\mathbf{F}(\mathbf{y})\|^2 - 2\mathbf{y}^{\mathrm{T}}\mathbf{F}^-(\mathbf{y}) + 2\sigma^2 \mathrm{div}\left\{\mathbf{F}^-(\mathbf{y})\right\}\right) p(\mathbf{x},\mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\ &\quad + \mathscr{E}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} - N\sigma^2 \\ &= \int_{\mathbb{R}^N} \left(\|\mathbf{F}(\mathbf{y})\|^2 - 2\mathbf{y}^{\mathrm{T}}\mathbf{F}^-(\mathbf{y}) + 2\sigma^2 \mathrm{div}\left\{\mathbf{F}^-(\mathbf{y})\right\}\right) r(\mathbf{y})\mathrm{d}\mathbf{y} \\ &\quad + \mathscr{E}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} - N\sigma^2 \\ &= \sum_{n=1}^{N} \Big(\int_{\mathbb{R}^N} f_n^2(\mathbf{y}) r(\mathbf{y})\mathrm{d}\mathbf{y} - 2 \underbrace{\int_{\mathbb{R}^N} y_n f_n(\mathbf{y}-\mathbf{e}_n) r(\mathbf{y})\mathrm{d}\mathbf{y}}_{\text{(I)}} \\ &\quad + 2\sigma^2 \underbrace{\int_{\mathbb{R}^N} \frac{\partial f_n(\mathbf{y}-\mathbf{e}_n)}{\partial y_n} r(\mathbf{y})\mathrm{d}\mathbf{y}}_{\text{(II)}} \Big) + \mathscr{E}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} - N\sigma^2 \end{aligned}$$

(I) Using a change of variable, (I) can be rewritten as:

$$\int_{\mathbb{R}^N} y_n f_n(\mathbf{y}-\mathbf{e}_n) r(\mathbf{y})\mathrm{d}\mathbf{y} = \int_{\mathbb{R}^N} (y_n+1) f_n(\mathbf{y}) r(\mathbf{y}+\mathbf{e}_n)\mathrm{d}\mathbf{y}$$

(II) Using an integration by parts followed by a change of variable, (II) can be rewritten as:

$$\int_{\mathbb{R}^N} \frac{\partial f_n(\mathbf{y}-\mathbf{e}_n)}{\partial y_n} r(\mathbf{y})\mathrm{d}\mathbf{y} = -\int_{\mathbb{R}^N} f_n(\mathbf{y}) \frac{\partial r(\mathbf{y}+\mathbf{e}_n)}{\partial y_n}\mathrm{d}\mathbf{y}$$

where we have assumed that $\lim_{\mathbf{y}\to\pm\boldsymbol{\infty}} f_n(\mathbf{y}-\mathbf{e}_n) r(\mathbf{y}) = 0$.

The expectation of PURE can be thus re-expressed as:

$$\begin{aligned} N\mathscr{E}\{\epsilon\} &= \sum_{n=1}^{N} \Bigg(\int_{\mathbb{R}^N} \Big(f_n^2(\mathbf{y}) r(\mathbf{y}) - 2(y_n+1) f_n(\mathbf{y}) r(\mathbf{y}+\mathbf{e}_n) \\ &\quad - 2\sigma^2 f_n(\mathbf{y}) \frac{\partial r(\mathbf{y}+\mathbf{e}_n)}{\partial y_n} \Big)\mathrm{d}\mathbf{y} \Bigg) + \mathscr{E}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} - N\sigma^2 \end{aligned}$$

Using calculus of variations, we find that the minimizer of the expected PURE/MSE is finally given by

$$f_n^{\mathrm{opt}}(\mathbf{y}) = (y_n+1)\frac{r(\mathbf{y}+\mathbf{e}_n)}{r(\mathbf{y})} + \frac{\sigma^2}{r(\mathbf{y})}\frac{\partial r(\mathbf{y}+\mathbf{e}_n)}{\partial y_n}$$

■

As in the AWGN case, the evaluation of the PDF of the noisy data is required to implement the MMSE estimator of Poisson intensities. The two strategies evoked in Section 2.5 (i.e. empirical estimation from the data or prior assumption on the noise-free data) are still conceivable, but both of them are challenging. Instead, we propose to take advantage of the quadratic form of PURE to build our Poisson intensity estimator $\mathbf{F}$ as a linear expansion of K thresholds (LET), as in the SURE-LET case (2.19). Consequently, the search for the optimal (in the minimum PURE sense) vector of parameters $\mathbf{a} = [a_1, a_2, \ldots, a_k]^{\mathrm{T}}$ boils down to the solution of the following *linear* system of K equations:

$$\begin{aligned} \sum_{l=1}^{K} \underbrace{\mathbf{F}_k(\mathbf{y})^{\mathrm{T}}\mathbf{F}_l(\mathbf{y})}_{[\mathbf{M}]_{k,l}} a_l &= \underbrace{\mathbf{y}^{\mathrm{T}}\mathbf{F}_k^-(\mathbf{y}) - \sigma^2 \mathrm{div}\left\{\mathbf{F}_k^-(\mathbf{y})\right\}}_{[\mathbf{c}]_k} \quad \text{for } k = 1, 2, \ldots K \\ &\Updownarrow \\ \mathbf{Ma} &= \mathbf{c} \end{aligned} \tag{6.13}$$

6.6 Summary

In this chapter, we have shown that it is possible to derive an unbiased estimate of the MSE beyond the restrictive case of AWGN considered in the original SURE theory. More generally, we have devised a non-Bayesian procedure (PURE-LET) to estimate Poisson intensities degraded by AWGN.

Although an optimal MMSE estimator can be theoretically found, its implementation is hardly feasible in practice. Similarly to the AWGN, we have thus proposed to use the Poisson's unbiased risk estimate (PURE) to optimize a linear expansion of thresholds (LET). Thanks to the quadratic form of PURE, this optimization boils down to the resolution of a linear system of equations, which ensures a low computational burden.

Chapter 7

PURE-LET in the Unnormalized Haar Wavelet Representation

7.1 Motivations

While an independent additive white Gaussian noise (AWGN) model is easily characterized in an orthonormal wavelet representation (see Chapter 3), the Poisson statistics are much more difficult to tract in the wavelet domain, due to their signal-dependent nature. An independent denoising of each wavelet subband, which ensures a fast algorithm, thus seems unachievable. However, thanks to a fundamental property of the Poisson distribution ("conservation under summation"), the oldest wavelet transform comes back in the forefront. Indeed, the use of the (unnormalized) *Haar* wavelet transform, which yields nothing but sums (resp. differences) of Poisson distributions in its lowpass (resp. highpass) channel, makes it possible to derive independent Poisson unbiased risk estimate (PURE) for each wavelet subband. From a practical point of view, the unnormalized Haar wavelet transform can thus be considered as the most suitable candidate for the development of a fast PURE-LET algorithm[1].

7.2 The Unnormalized Haar WaveletTransform

The unnormalized Haar discrete wavelet transform (DWT) can be seen as a standard two-channel filterbank (see Figure 7.1). The analysis pair of lowpass/highpass filters

[1]This chapter is an extension to the work presented in our recently published paper [165].

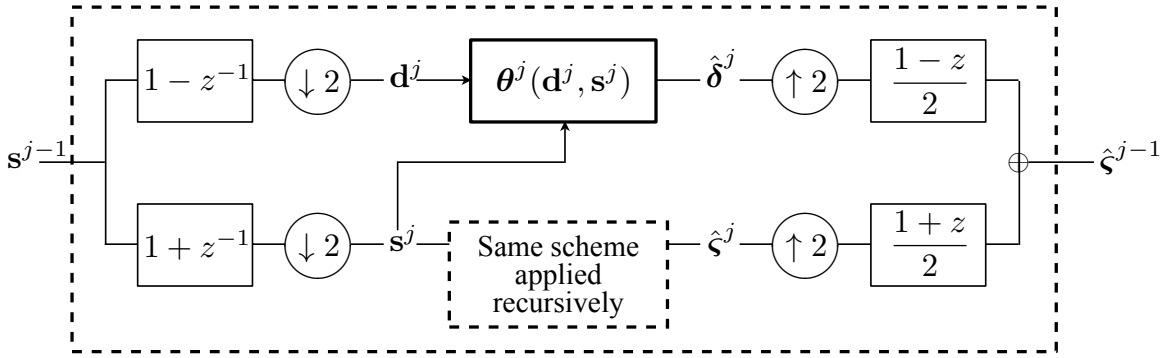


Figure 7.1: Filter bank implementation of the unnormalized Haar discrete wavelet transform and principle of the class of denoising algorithms described by (7.1). The scheme is applied recursively on the low-pass channel output.

is given in the z-transform domain by $H_a(z) = 1 + z^{-1}$, $G_a(z) = 1 - z^{-1}$ and the corresponding synthesis pair is $H_s(z) = H_a(z^{-1})/2$, $G_s(z) = G_a(z^{-1})/2$.

In this chapter, the unnormalized Haar scaling coefficients of the measurement $\mathbf{y} = \mathbf{z} + \mathbf{b}$ at scales $j = 1, \ldots, J$ are denoted by $\mathbf{s}^j \in \mathbb{R}^{N_j}$, where $N_j = N/2^j$, and $\mathbf{d}^j \in \mathbb{R}^{N_j}$ stands for the associated wavelet coefficients (we assume that the signal dimension is divisible by 2^J). Setting $\mathbf{s}^0 = \mathbf{y}$, these coefficients are obtained from the following sums and differences:

$$\left\{ \begin{array}{ll} s_n^j & = s_{2n}^{j-1} + s_{2n-1}^{j-1}, \\ d_n^j & = s_{2n}^{j-1} - s_{2n-1}^{j-1}, \end{array} \right. \quad \text{for } j = 1, \ldots, J$$

The original sequence $\mathbf{y} = \mathbf{s}^0$ is simply recovered by computing

$$\left\{ \begin{array}{ll} s_{2n}^{j-1} & = (s_n^j + d_n^j)/2, \\ s_{2n-1}^{j-1} & = (s_n^j - d_n^j)/2, \end{array} \right. \quad \text{for } j = J, \ldots, 1$$

Similarly, we denote by $\boldsymbol{\varsigma}^j$ and $\boldsymbol{\delta}^j$ the scaling and wavelet coefficients of the original signal $\mathbf{x}$ at a given scale j. Note that, by linearity of the wavelet transform, we have $\mathscr{E}\left\{d_n^j\right\} = \delta_n^j$ and $\mathscr{E}\left\{s_n^j\right\} = \varsigma_n^j$.

The key properties of the unnormalized Haar DWT are the following.

1. It is an *orthogonal* transform. In particular, we can split the MSE into subband-specific error terms:

$$\text{MSE} = \frac{2^{-J}}{N} \left\| \hat{\boldsymbol{\varsigma}}^J - \boldsymbol{\varsigma}^J \right\|^2 + \sum_{j=1}^{J} \frac{2^{-j}}{N} \left\| \hat{\boldsymbol{\delta}}^j - \boldsymbol{\delta}^j \right\|^2$$

This implies that we can minimize the MSE for each subband independently, while ensuring a global signal-domain MSE minimization.

2. At a given scale j, the scaling coefficients of an input vector of independent Poisson random variables are also independent Poisson random variables, because the sum of independent Poisson random variables is also a Poisson random variable with intensity equal to the summed intensities.

7.3 Haar-Wavelet-Domain PURE

In principle, the estimate $\hat{\boldsymbol{\delta}}^j$ may be constructed using all available subbands of the measurement $\mathbf{y}$; in practice, however, standard thresholding techniques only use the corresponding wavelet subband of the measurement, $\mathbf{d}^j$. In the sequel, we will consider denoising algorithms where $\hat{\boldsymbol{\delta}}^j$ also depends on the scaling coefficients at the same scale $\mathbf{s}^j$. This means that we have the following functional relationship:

$$\hat{\boldsymbol{\delta}}^j = \boldsymbol{\theta}^j(\mathbf{d}^j, \mathbf{s}^j). \tag{7.1}$$

As usual, the lowpass residual is not processed, i.e. $\hat{\boldsymbol{\varsigma}}^J = \mathbf{s}^J$. Our algorithmic framework is illustrated in Figure 7.1.

The above choice is advantageous from a computational standpoint because such a restoration procedure can be implemented in parallel with the wavelet decomposition, which yields the scaling coefficients $\mathbf{s}^j$ as by-products. Furthermore, this framework comprises advanced denoising schemes that exploit interscale dependencies via the scaling coefficients $\mathbf{s}^j$ (recall Section 3.3 and see Section 7.5.2). Finally, it allows us to minimize $\mathrm{MSE}_j = \|\boldsymbol{\theta}^j(\mathbf{d}^j, \mathbf{s}^j) - \boldsymbol{\delta}^j\|^2/N_j$ independently for each wavelet subband. We will thus focus on a fixed scale and drop the superscript j to simplify the notations.

In the Gaussian case, the scaling coefficients $\mathbf{s}^j$ are statistically independent of the wavelet coefficients $\mathbf{d}^j$ (in an orthogonal wavelet domain). This is in contrast with the Poisson case, for which $\mathbf{d}^j$ and $\mathbf{s}^j$ are statistically dependent and even correlated. This makes the derivation of a bivariate (i.e. involving both $\mathbf{d}^j$ and $\mathbf{s}^j$) SURE-like MSE estimate less obvious. In the following theorem, we give an expression of an unbiased estimate of the MSE in a given subband j of the unnormalized Haar DWT defined in Section 7.2. This result serves as a data-dependent quantitative measure to be minimized for Poisson intensity estimation in AWGN.

We recall that the family of vectors $(\mathbf{e}_n)_{n=1,\ldots,N_j}$ denotes the canonical basis of $\mathbb{R}^{N_j}$, i.e. all components of $\mathbf{e}_n$ are zero, except for the n^{th} component, which is equal to one.

Theorem 6. *Let $\boldsymbol{\theta}(\mathbf{d},\mathbf{s}) = \boldsymbol{\theta}^j(\mathbf{d}^j,\mathbf{s}^j)$ be an estimate of the noise-free wavelet coefficients $\boldsymbol{\delta} = \boldsymbol{\delta}^j$. Define $\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})$ and $\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})$ by*

$$\begin{cases} \theta_n^+(\mathbf{d},\mathbf{s}) = \theta_n(\mathbf{d}+\mathbf{e}_n, \mathbf{s}-\mathbf{e}_n) \\ \theta_n^-(\mathbf{d},\mathbf{s}) = \theta_n(\mathbf{d}-\mathbf{e}_n, \mathbf{s}-\mathbf{e}_n) \end{cases} \tag{7.2}$$

Then the random variable

$$\begin{aligned} \epsilon_j \;=\; & \frac{1}{N_j}\Big(\|\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\|^2 + \|\mathbf{d}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{s} - N_j\sigma_j^2 \\ & -\mathbf{d}^{\mathrm{T}}\big(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) + \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\big) - \mathbf{s}^{\mathrm{T}}\big(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) - \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\big) \\ & +\sigma_j^2\Big(\mathrm{div}_{\mathbf{d}}\left\{\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) + \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right\} + \mathrm{div}_{\mathbf{s}}\left\{\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) - \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right\}\Big)\Big) \end{aligned} \tag{7.3}$$

is an unbiased *estimate of the expected MSE for the subband under consideration, i.e., $\mathscr{E}\{\epsilon_j\} = \mathscr{E}\{MSE_j\}$.*

Proof. The proof relies centrally on the fact that, within a given scale, the scaling coefficients are independent Poisson random variables degraded by AWGN of variance $\sigma_j^2 = 2^j\sigma^2$. We consider the case where $j=1$, so that we can use $\mathbf{y} = \mathbf{s}^{j-1}$ and $\mathbf{x} = \boldsymbol{\varsigma}^{j-1}$ to avoid superscripts.

We first develop the squared error between $\boldsymbol{\delta}$ and its estimate $\boldsymbol{\theta}(\mathbf{d},\mathbf{s})$, using the fact that $\boldsymbol{\delta}$ is a deterministic quantity:

$$\mathscr{E}\left\{\|\boldsymbol{\theta}(\mathbf{d},\mathbf{s}) - \boldsymbol{\delta}\|^2\right\} = \mathscr{E}\left\{\|\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\|^2\right\} + \underbrace{\|\boldsymbol{\delta}\|^2}_{\text{(I)}} - 2\underbrace{\mathscr{E}\left\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\right\}}_{\text{(II)}} \tag{7.4}$$

Now, we evaluate the two expressions (I) and (II) that involve the unknown data $\boldsymbol{\delta}$.

(I) $\|\boldsymbol{\delta}\|^2 = \sum_{n=1}^{N_1} (\delta_n)^2$: We notice that

$$(\delta_n)^2 = \mathscr{E}\left\{x_{2n}(y_{2n} - y_{2n-1})\right\} + \mathscr{E}\left\{x_{2n-1}(y_{2n-1} - y_{2n})\right\}$$

By applying Lemma 4 and Lemma 1 for $\theta(y) = y - y_{2n-1}$ and for $\theta(y) = y - y_{2n}$, by using the fact that y_{2n} and y_{2n-1} are statistically independent, we get

$$(\delta_n)^2 = \mathscr{E}\left\{(y_{2n} - y_{2n-1})^2 - (y_{2n} + y_{2n-1})\right\} - 2\sigma^2 = \mathscr{E}\left\{d_n^2 - s_n\right\} - \sigma_1^2$$

Therefore $\|\boldsymbol{\delta}\|^2 = \mathscr{E}\left\{\|\mathbf{d}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{s}\right\} - N_1\sigma_1^2$

(II) $\mathscr{E}\left\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\right\} = \sum_{n=1}^{N_1} \mathscr{E}\left\{\delta_n\theta_n(\mathbf{d},\mathbf{s})\right\}$: We can successively write

$$\begin{aligned}
&\mathscr{E}\left\{\delta_n\theta_n(\mathbf{d},\mathbf{s})\right\} \\
&\quad = \quad \mathscr{E}\left\{x_{2n}\theta_n(\mathbf{d},\mathbf{s})\right\} - \mathscr{E}\left\{x_{2n-1}\theta_n(\mathbf{d},\mathbf{s})\right\} \\
&\quad \overset{\text{Lemma 4}}{=} \quad \mathscr{E}\left\{y_{2n}\theta_n(\mathbf{d}-\mathbf{e}_n,\mathbf{s}-\mathbf{e}_n)\right\} - \mathscr{E}\left\{y_{2n-1}\theta_n(\mathbf{d}+\mathbf{e}_n,\mathbf{s}-\mathbf{e}_n)\right\} - \\
&\qquad\qquad \mathscr{E}\left\{b_{2n}\theta_n(\mathbf{d}-\mathbf{e}_n,\mathbf{s}-\mathbf{e}_n)\right\} + \mathscr{E}\left\{b_{2n-1}\theta_n(\mathbf{d}+\mathbf{e}_n,\mathbf{s}-\mathbf{e}_n)\right\} \\
&\quad \overset{\text{Lemma 1}}{=} \quad \mathscr{E}\left\{\frac{y_{2n}-y_{2n-1}}{2}\left(\theta_n^-(\mathbf{d},\mathbf{s})+\theta_n^+(\mathbf{d},\mathbf{s})\right)\right\} + \\
&\qquad\qquad \mathscr{E}\left\{\frac{y_{2n}+y_{2n-1}}{2}\left(\theta_n^-(\mathbf{d},\mathbf{s})-\theta_n^+(\mathbf{d},\mathbf{s})\right)\right\} - \\
&\qquad\qquad \sigma^2\left(\mathscr{E}\left\{\frac{\partial\theta_n^-(\mathbf{d},\mathbf{s})}{\partial d_n}+\frac{\partial\theta_n^-(\mathbf{d},\mathbf{s})}{\partial s_n}\right\}+\mathscr{E}\left\{\frac{\partial\theta_n^+(\mathbf{d},\mathbf{s})}{\partial d_n}-\frac{\partial\theta_n^+(\mathbf{d},\mathbf{s})}{\partial s_n}\right\}\right) \\
&\quad = \quad \frac{1}{2}\mathscr{E}\left\{d_n\left(\theta_n^-(\mathbf{d},\mathbf{s})+\theta_n^+(\mathbf{d},\mathbf{s})\right)+s_n\left(\theta_n^-(\mathbf{d},\mathbf{s})-\theta_n^+(\mathbf{d},\mathbf{s})\right)\right\} - \\
&\qquad\qquad \sigma^2\mathscr{E}\left\{\frac{\partial}{\partial d_n}\left(\theta_n^-(\mathbf{d},\mathbf{s})+\theta_n^+(\mathbf{d},\mathbf{s})\right)+\frac{\partial}{\partial s_n}\left(\theta_n^-(\mathbf{d},\mathbf{s})-\theta_n^+(\mathbf{d},\mathbf{s})\right)\right\}
\end{aligned}$$

and thus, since $\sigma_1^2 = 2\sigma^2$,

$$\begin{aligned}
2\mathscr{E}\left\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\right\} = {} & \mathscr{E}\left\{\mathbf{d}^{\mathrm{T}}\left(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})+\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right)+\mathbf{s}^{\mathrm{T}}\left(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})-\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right)\right\} + \\
& \sigma_1^2\mathscr{E}\left\{\mathrm{div}_{\mathbf{d}}\left\{\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})+\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right\}+\mathrm{div}_{\mathbf{s}}\left\{\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})-\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right\}\right\}
\end{aligned}$$

Putting these two results (I),(II) back into (7.4) gives the desired equality. For $j \geq 2$, the proof is based on the same idea. ■

Recently, a result similar to Equ. (7.3) was independently derived by Hirakawa *et. al.* in [166] for the pure Poisson case ($\sigma = 0$).

7.4 Haar-Wavelet-Domain PUREshrink

As a benchmark for illustrating our approach, we propose a wavelet-domain estimator which consists in applying the popular *soft-threshold* (ST) with a PURE-optimized threshold. Our *PUREshrink* estimator can be viewed as the transposition of Donoho and Johnstone's Gaussian *SUREshrink* [71] to Poisson noise removal. An important difference is that the method described in [71] forces the threshold to be smaller than the universal threshold (otherwise it is set to the value of the universal threshold); this is known to be suboptimal for image-denoising applications. Our threshold optimization totally relies on the minimization of PURE (without restrictions).

Contrary to the Gaussian case, where the noise is stationary and completely described by its variance[2], for Poisson data, the amount of noise directly depends on the intensity we want to estimate. In the unnormalized Haar wavelet-domain, each wavelet coefficient follows a Skellam distribution [167], whose variance is equal to the sum of the two underlying Poisson intensities. By further considering the variance induced by the AWGN, the overall variance of each wavelet coefficient d_n^j is given by $\mathrm{var}\left\{d_n^j\right\} = \varsigma_n^j + \sigma_j^2$, which can be roughly approximated by $\mathrm{var}\left\{d_n^j\right\} \simeq s_n^j + \sigma_j^2$. Note that in the *Haar-Fisz* transform [156], devised to stabilize the variance of a pure Poisson process, the scaling coefficients are also considered as an estimate of the local noise variance. The *PUREshrink* estimator can therefore be defined as

$$\theta_n^{\mathrm{ST}}(\mathbf{d}, \mathbf{s}; a) = \mathrm{sign}(d_n)\, \max(|d_n| - a\sqrt{|s_n| + \sigma^2}, 0), \tag{7.5}$$

where, for each wavelet subband, the parameter a is set to the value that minimizes the PURE (7.3) with $\boldsymbol{\theta}(\mathbf{d}, \mathbf{s}) = \boldsymbol{\theta}^{\mathrm{ST}}(\mathbf{d}, \mathbf{s}; a)$. As observed in Figure 7.2, the minimum of the PURE curve for various values of a closely coincides with the minimum of the MSE curve, confirming the reliability of PURE as an estimate of the MSE.

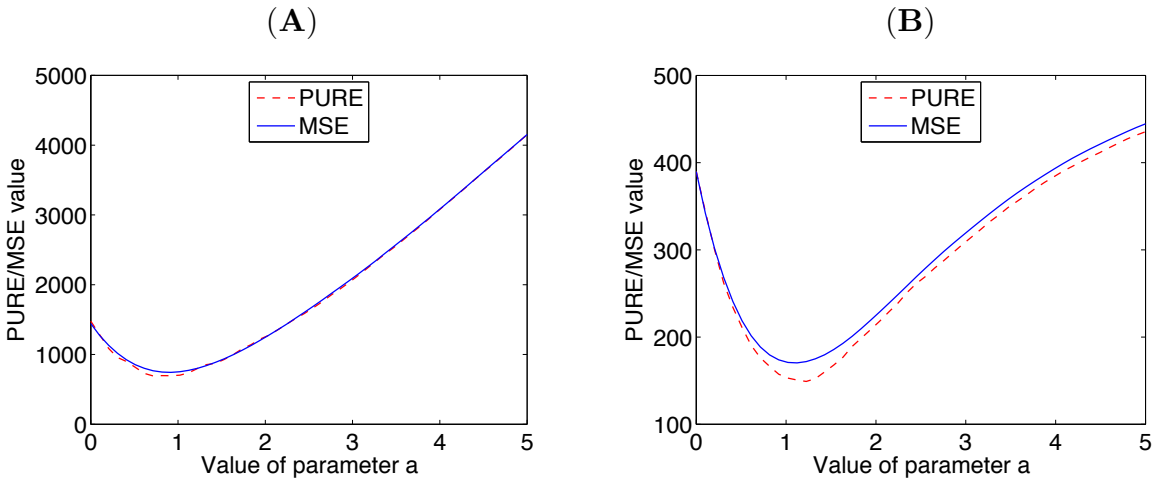


Figure 7.2: Reliability of PURE as an estimate of the MSE. (**A**) For a 32×32 Haar wavelet subband of the *Cameraman* image at PSNR $= 15.89\,\mathrm{dB}$. (**B**) For a 64×64 Haar wavelet subband of the *Boat* image at PSNR $= 12.18\,\mathrm{dB}$.

[2]The noise is usually assumed to be zero-mean.

7.5 Haar-Wavelet-Domain PURE-LET

Following the *PURE-LET* strategy introduced in Section 6.5, we propose to consider a wavelet estimator that is formulated as a linear expansion of K thresholds (LET):

$$\boldsymbol{\theta}^{\mathrm{LET}}(\mathbf{d},\mathbf{s};\mathbf{a}) = \sum_{k=1}^{K} a_k \boldsymbol{\theta}_k(\mathbf{d},\mathbf{s}),$$

where the $\boldsymbol{\theta}_k$'s are generic estimators that will be specified later in this section. Thanks to this linear parameterization, the unbiased estimate of the MSE given in (7.3) is *quadratic* with respect to the parameters $\mathbf{a} \in \mathbb{R}^K$. Therefore, its minimization boils down to the resolution of a *linear* system of equations with *small* dimension K:

$$\mathbf{a} = \mathbf{M}^{-1}\mathbf{c} \tag{7.6}$$

where for $1 \leq k, l \leq K$,

$$\begin{cases} M_{k,l} &= \boldsymbol{\theta}_k(\mathbf{d},\mathbf{s})^{\mathrm{T}} \boldsymbol{\theta}_l(\mathbf{d},\mathbf{s}) \\ c_k &= \frac{1}{2}\Big(\mathbf{d}^{\mathrm{T}}(\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s}) + \boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s})) + \mathbf{s}^{\mathrm{T}}(\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s}) - \boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s}))\Big) - \\ & \quad \frac{\sigma^2}{2}\Big(\mathrm{div}_{\mathbf{d}}\left\{\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) + \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right\} + \mathrm{div}_{\mathbf{s}}\left\{\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) - \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right\}\Big) \end{cases} \tag{7.7}$$

The definition of $\boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s})$ and $\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s})$ is similar to (7.2).

7.5.1 Pointwise Thresholding

Similarly to the AWGN case, we propose a linearly-parameterized thresholding function with $K = 2$ parameters (a_1 and a_2), whose n^{th} component is defined by

$$\theta_n^{\mathrm{LET}_0}(\mathbf{d},\mathbf{s};[a_1\, a_2]^{\mathrm{T}}) = a_1 d_n + a_2 \left(1 - \exp\left(-\frac{d_n^2}{2T_n^2}\right)\right) d_n \tag{7.8}$$

In this expression, the linear parameters a_1 and a_2 define a compromise between two regimes: either the wavelet coefficient d_n is kept as is (signal preservation) or it is shrinked towards zero (noise suppression). The exponential function has the advantage of being smooth, which reduces the variance of the estimator.

As in the *PUREshrink* estimator, the threshold T_n is directly linked to the local noise variance, estimated from the magnitude of the corresponding scaling coefficient $|s_n|$. However, thanks to the degrees of freedom provided by the two *linear* parameters a_1 and a_2, the value of this threshold does not need to be optimized. By running several simulations, we found that $T_n^2 = 6(|s_n| + \sigma^2)$ constituted a good choice, inducing no significant loss compared to a subband-optimized threshold. Our experiments (see Figure 7.4) show that the above thresholding function is already more efficient (approx. $+0.25\,\mathrm{dB}$) than the previously presented *PUREshrink* (7.5).

7.5.2 Interscale Estimator

Interscale sign dependencies

The integration of interscale dependencies has already been shown to bring a substantial improvement in the context of AWGN removal (Section 3.3). Therefore, we propose here an analogous interscale wavelet thresholding, but for Poisson intensity estimation. The idea is to exploit the scaling coefficients $\mathbf{s}$ to "predict" and reinforce the significant wavelet coefficients of $\mathbf{d}$ at the same scale. Indeed, the scaling coefficients offer improved SNR because they arise from Poisson random variables with summed intensities. They also contain all the information about coarser-scale wavelet coefficients.

To construct an interscale predictor of the wavelet coefficient d_n, we simply compute the difference between the two scaling coefficients that surround s_n:

$$\tilde{d}_n = s_{n-1} - s_{n+1}$$

Indeed, the lowpass and highpass *Haar* filters have the same group delay. The whole procedure thus boils down to applying a centered gradient filter on the lowpass subband, as illustrated in the diagram of Figure 7.3.

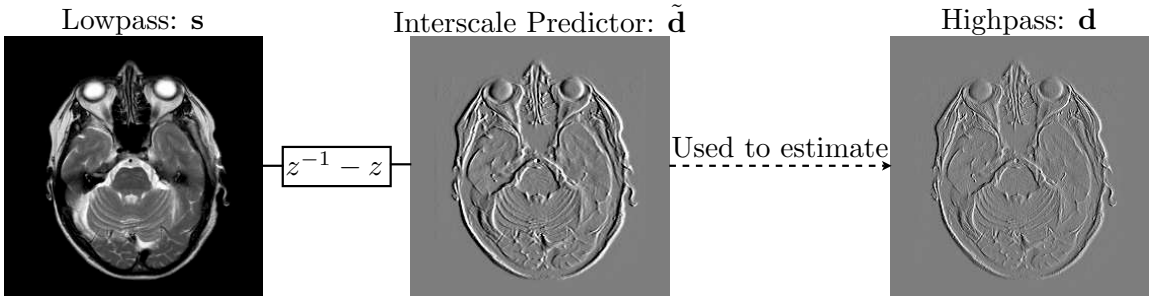


Figure 7.3: The interscale predictor of subband HL_j (resp. LH_j, resp. HH_j) is obtained by horizontally (resp. vertically, resp. horizontally and vertically) filtering the same-scale lowpass subband LL_j with the centered gradient filter, whose z-transform is $z^{-1} - z$.

By taking a closer look at Figure 7.3, it can be observed that the signs of the interscale predictor coefficients are consistent with those of the corresponding highpass subband. This suggests adding a term proportional to this interscale predictor into the simple thresholding function (7.8), leading to

$$\theta_n^{\mathrm{LET}_1}(\mathbf{d}, \mathbf{s}; [a_1\, a_2\, a_3]^{\mathrm{T}}) = \theta_n^{\mathrm{LET}_0}(\mathbf{d}, \mathbf{s}; [a_1\, a_2]^{\mathrm{T}}) + a_3 \underbrace{(s_{n-1} - s_{n+1})}_{\tilde{d}_n} \tag{7.9}$$

This simple strategy brings significant improvements (approximately +0.5 dB). This was confirmed by multiple experiments on standard grayscale images; some of these simulations are reported in Figure 7.4.

Interscale amplitude-sign dependencies

Further improvements can be obtained by grouping together wavelet coefficients of similar magnitudes (Section 3.3). To increase the robustness w.r.t. noise, this grouping is based on the magnitude of a smoothed version $\mathbf{p}$ of the previously defined interscale predictor $\tilde{\mathbf{d}}$. The smoothed version of the interscale predictor is simply obtained by applying a normalized Gaussian kernel on the absolute value of $\tilde{\mathbf{d}}$, i.e., $p_n = \sum_k |\tilde{d}_k| f_{n-k}$, where $f_k = e^{-k^2/2}/\sqrt{2\pi}$. The proposed interscale wavelet thresholding is thus finally defined as

$$\begin{aligned}\theta_n^{\mathrm{LET}_2}(\mathbf{d},\mathbf{s};\mathbf{a},\mathbf{b}) \\ = \quad \exp\left(-\frac{p_n^2}{2T_n^2}\right)\theta_n^{\mathrm{LET}_1}(\mathbf{d},\mathbf{s};\mathbf{a}) + \left(1-\exp\left(-\frac{p_n^2}{2T_n^2}\right)\right)\theta_n^{\mathrm{LET}_1}(\mathbf{d},\mathbf{s};\mathbf{b})\end{aligned} \tag{7.10}$$

where $T_n^2 = 6(|s_n| + \sigma^2)$ is the same threshold as the one that appears in LET_0 (7.8).

The above thresholding function has mainly two regimes: when the squared magnitude of the predictor p_n is small with respect to the local noise variance, LET_2 essentially behaves like LET_1 with parameter $\mathbf{a}$; when p_n^2 is large with respect to the local noise variance, LET_2 essentially behaves like LET_1 with parameter $\mathbf{b}$. This "classification" based on the predictor coefficients increases the adaptivity of the denoising process.

In Figure 7.4 and 7.5, it is seen that this latter interscale wavelet estimator clearly gives the best results, both quantitatively and visually, among all estimators presented in this section. Note that the PURE-based adjustment of the parameters $\mathbf{a}$ and $\mathbf{b}$ gives a SNR gain that is very close to the optimum (which is obtained from an *oracle* adjustment of these parameters using the knowledge of the original image).

7.5.3 Multichannel Estimator

In this section, we propose a multichannel extension of the interscale Haar wavelet estimator (7.10). According to the notations introduced in Section 3.4, Equ. (3.19), the MSE inside each $C \times N_j$ Haar wavelet subband j is defined as:

$$\mathrm{MSE}_j = \frac{1}{CN_j}\mathrm{trace}\left\{(\boldsymbol{\theta}^j(\mathbf{d}^j,\mathbf{s}^j) - \boldsymbol{\delta}^j)(\boldsymbol{\theta}^j(\mathbf{d}^j,\mathbf{s}^j) - \boldsymbol{\delta}^j)^{\mathrm{T}}\right\} \tag{7.11}$$

Multichannel Haar-Wavelet-Domain PURE

We assume that the noisy measurements $y_{c,n}$ follow i.i.d. Poisson laws degraded by i.i.d. AWGNs of variance σ^2. In this framework, we can state the following result:

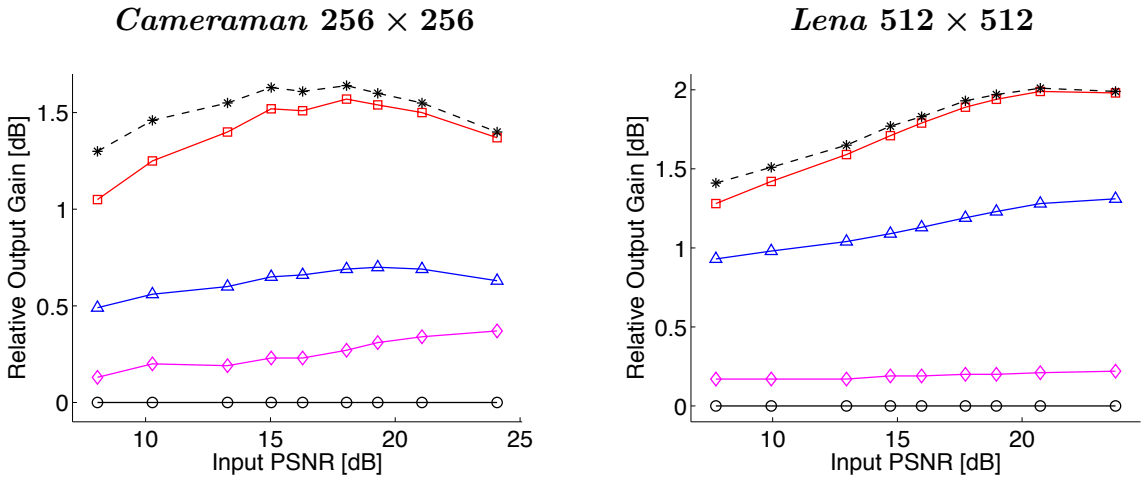


Figure 7.4: PSNR improvements brought by the *PURE-LET* strategy over *PUREshrink* (Formula (7.5): "○" markers), for two standard grayscale images. Formula (7.8): "◇" markers. Formula (7.9): "△" markers. Formula (7.10): "□" markers). Oracle results for Formula (7.10) are shown with "∗" markers.

Corollary 7. *Let* $\boldsymbol{\theta}(\mathbf{d},\mathbf{s}) = \boldsymbol{\theta}^j(\mathbf{d}^j,\mathbf{s}^j)$ *be an estimate of the noise-free wavelet coefficients* $\boldsymbol{\delta} = \boldsymbol{\delta}^j$ *and let* $\mathbf{e}_{c,n}$ *denote a* $C \times N$ *matrix filled with zeros, except at the position* (c,n) *which is set to one. Define* $\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})$ *and* $\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})$ *by*

$$\begin{cases} \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s}) = [\theta_{c,n}(\mathbf{d}+\mathbf{e}_{c,n},\mathbf{s}-\mathbf{e}_{c,n})]_{1\le c\le C, 1\le n\le N_j} \\ \boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) = [\theta_{c,n}(\mathbf{d}-\mathbf{e}_{c,n},\mathbf{s}-\mathbf{e}_{c,n})]_{1\le c\le C, 1\le n\le N_j} \end{cases} \tag{7.12}$$

Then the random variable

$$\begin{aligned} \epsilon_j \;=\; & \frac{1}{CN_j}\mathrm{trace}\left\{\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\boldsymbol{\theta}(\mathbf{d},\mathbf{s})^{\mathrm{T}} + \mathbf{d}\mathbf{d}^{\mathrm{T}} - \mathbf{1}\mathbf{s}^{\mathrm{T}}\right\} - \sigma^2 - \\ & \frac{1}{CN_j}\mathrm{trace}\left\{\mathbf{d}\big(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) + \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\big)^{\mathrm{T}} + \mathbf{s}\big(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) - \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\big)^{\mathrm{T}}\right\} + \\ & \frac{\sigma_j^2}{CN_j}\Big(\mathrm{div}_{\mathbf{d}}\left\{\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) + \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right\} + \mathrm{div}_{\mathbf{s}}\left\{\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s}) - \boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\right\}\Big) \end{aligned} \tag{7.13}$$

is an unbiased *estimate of the MSE* (7.11) *for the subband under consideration, i.e.* $\mathscr{E}\{\epsilon_j\} = \mathscr{E}\{MSE_j\}$.

Here, $\mathrm{div}_{\mathbf{d}}\{\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\} = \sum_{c,n}\dfrac{\partial\theta_{c,n}(\mathbf{d},\mathbf{s})}{\partial d_{c,n}}$ and $\mathrm{div}_{\mathbf{s}}\{\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\} = \sum_{c,n}\dfrac{\partial\theta_{c,n}(\mathbf{d},\mathbf{s})}{\partial s_{c,n}}$ are two generalized divergence operators. The result provided in Corollary 7 directly comes from a natural vector/matrix extension of Theorem 6.

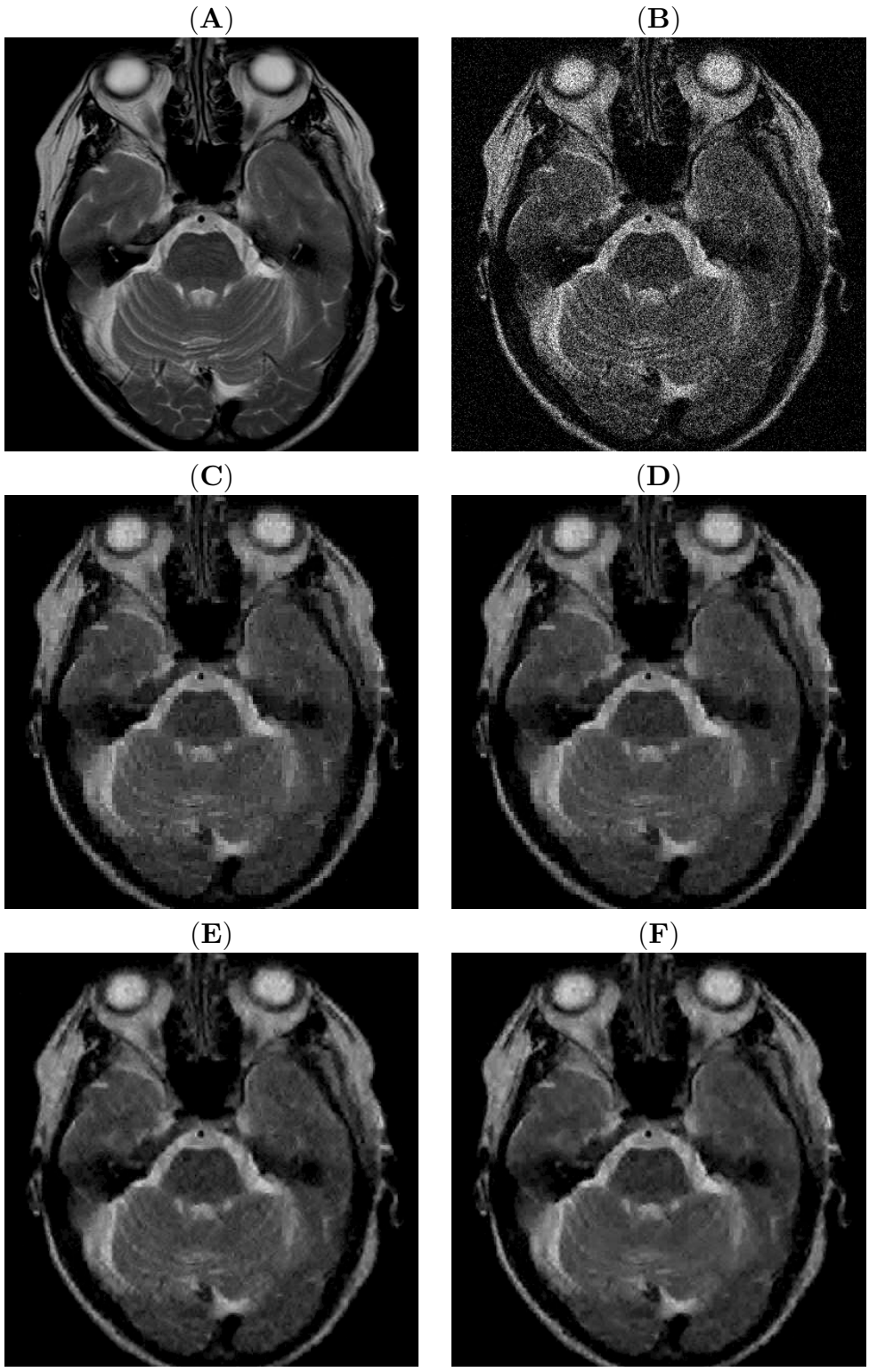


Figure 7.5: (**A**) Part of the original *MRI* slice ($I_{\max} = 20$). (**B**) Noisy realization of it: PSNR = 17.99 dB. (**C**) Denoised by *PUREshrink* (7.5): PSNR = 27.83 dB. (**D**) Denoised by *PURE-LET*$_0$ (7.8): PSNR = 28.03 dB. (**E**) Denoised by *PURE-LET*$_1$ (7.9): PSNR = 29.17 dB. (**F**) Denoised by *PURE-LET*$_2$ (7.10): PSNR = 29.73 dB.

Multichannel Haar-Wavelet-Domain PURE-LET

In the PURE-LET framework, the proposed multichannel Haar-wavelet-domain estimator is built as a linear expansion of K thresholds, i.e.

$$\boldsymbol{\theta}(\mathbf{d},\mathbf{s}) \;=\; \underbrace{[\mathbf{a}_1^{\mathrm{T}}\;\mathbf{a}_2^{\mathrm{T}}\ldots\mathbf{a}_K^{\mathrm{T}}]}_{\mathbf{A}^{\mathrm{T}}} \begin{bmatrix} \boldsymbol{\theta}_1(\mathbf{d},\mathbf{s}) \\ \boldsymbol{\theta}_2(\mathbf{d},\mathbf{s}) \\ \vdots \\ \boldsymbol{\theta}_K(\mathbf{d},\mathbf{s}) \end{bmatrix} \tag{7.14}$$

Thanks to this linear parameterization, the optimal set of $KC \times C$ parameters $\mathbf{A}$ (i.e. the minimizer of (7.13)) is the solution of a *linear* system of equations:

$$\mathbf{A}_{\mathrm{opt}} = \mathbf{M}^{-1}\mathbf{C} \tag{7.15}$$

where for $1 \le k,l \le K$

$$\begin{cases} \mathbf{M}_{k,l} \;=\; \boldsymbol{\theta}_k(\mathbf{d},\mathbf{s})\boldsymbol{\theta}_l(\mathbf{d},\mathbf{s})^{\mathrm{T}} \\ \mathbf{C}_k \;=\; \frac{1}{2}\left(\mathbf{d}\big(\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s})+\boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s})\big)^{\mathrm{T}} + \mathbf{s}\big(\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s})-\boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s})\big)^{\mathrm{T}}\right) - \\ \qquad \frac{\sigma^2}{2}\left(\mathrm{div}_1\left\{\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s})+\boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s})\right\} + \mathrm{div}_2\left\{\boldsymbol{\theta}_k^-(\mathbf{d},\mathbf{s})-\boldsymbol{\theta}_k^+(\mathbf{d},\mathbf{s})\right\}\right) \end{cases} \tag{7.16}$$

Multichannel Haar-Wavelet-Domain Interscale Thresholding

Similarly to the multichannel denoising of AWGN (Section 3.4.3), we propose to use the following multichannel Haar-wavelet-domain interscale thresholding:

$$\begin{aligned} &\boldsymbol{\theta}_n(\mathbf{d},\mathbf{s}) \\ &= \underbrace{\gamma_n(\mathbf{p}_n^{\mathrm{T}}\mathbf{p}_n)\gamma_n(\mathbf{d}_n^{\mathrm{T}}\mathbf{d}_n)}_{\text{small predictors and small coefficients}}\mathbf{a}_1^{\mathrm{T}}\mathbf{d}_n + \underbrace{\overline{\gamma}_n(\mathbf{p}_n^{\mathrm{T}}\mathbf{p}_n)\gamma_n(\mathbf{d}_n^{\mathrm{T}}\mathbf{d}_n)}_{\text{large predictors and small coefficients}}\mathbf{a}_2^{\mathrm{T}}\mathbf{d}_n + \\ &\quad \underbrace{\gamma_n(\mathbf{p}_n^{\mathrm{T}}\mathbf{p}_n)\overline{\gamma}_n(\mathbf{d}_n^{\mathrm{T}}\mathbf{d}_n)}_{\text{small predictors and large coefficients}}\mathbf{a}_3^{\mathrm{T}}\mathbf{d}_n + \underbrace{\overline{\gamma}_n(\mathbf{p}_n^{\mathrm{T}}\mathbf{p}_n)\overline{\gamma}_n(\mathbf{d}_n^{\mathrm{T}}\mathbf{d}_n)}_{\text{large predictors and large coefficients}}\mathbf{a}_4^{\mathrm{T}}\mathbf{d}_n + \\ &\quad \underbrace{\gamma_n(\mathbf{p}_n^{\mathrm{T}}\mathbf{p}_n)\,\mathbf{a}_5^{\mathrm{T}}\tilde{\mathbf{d}}_n + \overline{\gamma}_n(\mathbf{p}_n^{\mathrm{T}}\mathbf{p}_n)\,\mathbf{a}_6^{\mathrm{T}}\tilde{\mathbf{d}}_n}_{\text{sign consistency enhancement}} \end{aligned} \tag{7.17}$$

where $\gamma_n(x) = \exp\left(-\frac{|x|}{2\sqrt{C}T_n^2}\right)$ and $\overline{\gamma}_n(x) = 1-\gamma_n(x)$ are the two complementary grouping functions. The signal-dependent threshold is straightforwardly extended from the monochannel case, i.e. $T_n^2 = 6|\mathbf{1}^{\mathrm{T}}\mathbf{s}_n| + C\sigma^2$. When $C=1$, we obviously recover the monochannel denoiser defined in Equ. (7.10).

7.6 Experiments

In this section, we propose to compare our Haar-Wavelet-Domain *PURE-LET* approach with three multiscale-based methods in simulated experiments:

- A variant of the Haar-Fisz algorithm described in [156]: the Haar-Fisz variance-stabilizing transform (VST) is followed by a translation-invariant (TI) *SURE-shrink* [71], obtained by averaging several cycle-spins (CS) of a non-redundant wavelet transform. Due to the non shift-invariance of the Haar-Fisz transform, some "external" cycle-spins are also performed to the whole algorithm (Haar-Fisz VST + TI-*SUREshrink* + Haar-Fisz inverse VST), as suggested in [156]. The use of a "smooth" wavelet is also recommended: we have thus considered Daubechies' *symlets* with eight vanishing moments (*sym8*) [39].

- A standard variance-stabilizing transform followed by a high-quality AWGN removal algorithm: as VST, we have retained the popular generalized Anscombe transform (GAT) [168], which can also stabilize Poisson processes embedded in AWGN. For the denoising part, we have considered Portilla *et. al. BLS-GSM* [81], applied both in a non-redundant wavelet representation (*sym8*) and in a full steerable pyramid (FSP).

- A state-of-the-art denoising algorithm specifically designed for Poisson intensity estimation: we have retained Willett and Nowak's *Platelet* approach[3] [153]. Their Poisson intensity estimation consists in optimizing (through a penalized likelihood) the coefficients of polynomials fitted to a recursive dyadic partition of the support of the Poisson intensity. The shift-invariance of their approach can be increased by performing several cycle-spins.

For all wavelet-based methods presented in this section, we use the same number of decomposition levels, i.e. 4 (resp. 5) for 256×256 (resp. 512×512) images. Standard periodic boundary conditions have been applied. The input peak signal-to-noise ratios (PSNR[4]) are adjusted by rescaling the original test images, from a maximum intensity $I_{\max}$ of 120 to a minimum of 1.

7.6.1 Poisson Noise Removal

In this section, we only consider the estimation of the intensity of a pure Poisson process, i.e. without further AWGN degradation.

Table 7.1 summarizes the PSNRs obtained by the various algorithms applied in a *non-redundant* framework. It can be observed that the PURE-based approach

[3]Matlab code available at: `http://nislab.ee.duke.edu/NISLab/Platelets.html`

[4]Defined as: $\text{PSNR} = 10 \log_{10} \frac{I_{\max}^2}{\text{MSE}}$, where $I_{\max}$ is the maximum intensity of the noise-free image.

clearly outperforms (over 1 dB, on average) the VST-based (standard or sophisticated) wavelet denoisers. Note that the retained basis (*sym8*) is smoother, and thus more suitable for image-restoration tasks, than the basic *Haar* wavelet that we use. Our solution also gives significantly better PSNRs (~ 1 dB) than the non-redundant version of the *Platelet* approach.

In Figure 7.6, we compare the various algorithms applied in *redundant* representations. For the *Haar-Fisz+SUREshrink* and *Platelet* algorithms, as well as for the proposed *PURE-LET*(7.10), near translation invariance is achieved by performing several cycle-spins, respectively 20 internal and 5 external[5], 10, and 10. The redundant *BLS-GSM* consists in applying the *BLS-GSM* in a full steerable pyramid. We observe that the *non-redundant* version of the *PURE-LET*(7.10), which serves as a benchmark in this comparison, is already competitive with the *redundant* variants of the VST-based denoisers. Nevertheless, it is usually outperformed by its redundant version and by the state-of-the-art *Platelet* algorithm.

Two important aspects are not reflected in these comparisons: the visual quality and the computation time. Some denoised images are given in Figure 7.7 and Table 7.2 gives more insights regarding the tradeoff between the degree of redundancy and the computation time of the various algorithms. The *Platelet* procedure achieves the best PSNRs when considering a high number (25) of cyclic shifts. However, these results are obtained at a prohibitive computational cost for a routine use in real applications. Cyclic shifts of our PURE-based approach also brings some gains (around 1 dB), despite the fact that an independent "shift-wise" PURE minimization is sub-optimal (as shown in Chapter 4 for the Gaussian case). This suggests that deriving a rigorous unbiased estimate of the MSE for redundant processing of Poisson data could bring further improvement; this aspect will be examined in the next chapter.

Finally, we also compared the proposed *PURE-LET* algorithm with the interscale *SURE-LET* strategy, specifically devised for Gaussian noise statistics (see Chapter 3). For a fair comparison, we used an adapted implementation of the interscale *SURE-LET* algorithm that involved the same number of parameters as the present *PURE-LET* method. We also considered the same wavelet transformation, i.e. OWT Haar, for both techniques. As can be seen in Figure 7.8, applying the *SURE-LET* strategy in the VST-domain is less efficient for small intensities (over 0.5 dB loss for intensities lower than 10). This can be attributed to the rigorous minimization of an estimate of the actual MSE that is performed by the proposed *PURE-LET* algorithm.

7.6.2 Poisson and Gaussian Noise Removal

Unfortunately, there are only few algorithms (e.g. [169]) specifically designed for the estimation of Poisson intensities in AWGN. In Figure 7.9, we thus propose to

[5]Due to the non-shift-invariance of the Haar-Fisz variance-stabilizing transform.

Table 7.1: Comparison of *non-redundant* multiscale Poisson noise removal algorithms

Image	***Peppers* 256 × 256**						
$\mathbf{I_{max}}$	120	60	30	20	10	5	1
Input PSNR	23.92	20.92	17.91	16.17	13.14	10.12	3.14
Haar-Fisz+SUREshrink (OWT *sym8*)	29.27	27.30	24.54	23.82	21.98	21.03	18.69
GAT+*BLS-GSM* (OWT *sym8*)	29.90	28.02	26.16	25.13	23.10	20.48	15.55
Platelet (1 CS)	29.07	27.44	25.73	24.92	23.23	21.57	18.17
PURE-LET(7.10) **(OWT *Haar*)**	**30.28**	**28.51**	**26.72**	**25.70**	**23.81**	**21.99**	**18.92**
Image	***Cameraman* 256 × 256**						
$\mathbf{I_{max}}$	120	60	30	20	10	5	1
Input PSNR	24.08	21.07	18.05	16.29	13.28	10.27	3.28
Haar-Fisz+SUREshrink (OWT *sym8*)	28.55	26.42	24.40	23.10	22.40	20.97	18.68
GAT+*BLS-GSM* (OWT *sym8*)	29.68	27.79	25.95	24.88	22.82	20.55	15.49
Platelet (1 CS)	28.29	26.79	25.44	24.60	23.24	21.49	18.70
PURE-LET(7.10) **(OWT *Haar*)**	**30.13**	**28.30**	**26.58**	**25.56**	**23.93**	**22.41**	**19.32**
Image	***MIT* 256 × 256**						
$\mathbf{I_{max}}$	120	60	30	20	10	5	1
Input PSNR	25.77	22.78	19.77	18.00	15.01	12.00	5.00
Haar-Fisz+SUREshrink (OWT *sym8*)	28.81	26.77	24.26	22.90	21.76	19.98	16.08
GAT+*BLS-GSM* (OWT *sym8*)	29.83	27.70	25.57	24.38	22.07	19.83	14.30
Platelet (1 CS)	28.25	26.44	24.70	23.47	21.65	20.18	16.24
PURE-LET(7.10) **(OWT *Haar*)**	**30.06**	**27.97**	**25.99**	**24.87**	**23.00**	**21.20**	**17.39**
Image	***Moon* 512 × 512**						
$\mathbf{I_{max}}$	120	60	30	20	10	5	1
Input PSNR	26.25	23.24	20.24	18.48	15.46	12.46	5.46
Haar-Fisz+SUREshrink (OWT *sym8*)	29.03	26.83	25.02	24.33	23.56	22.79	21.97
GAT+*BLS-GSM* (OWT *sym8*)	29.10	27.35	25.66	24.71	23.41	21.94	18.77
Platelet (1 CS)	27.16	26.01	25.05	24.60	23.96	23.63	22.97
PURE-LET(7.10) **(OWT *Haar*)**	**29.62**	**27.96**	**26.57**	**25.88**	**24.92**	**24.24**	**23.16**

Note: The output PSNRs have been averaged over ten noise realizations, except for the *Platelet* approach.

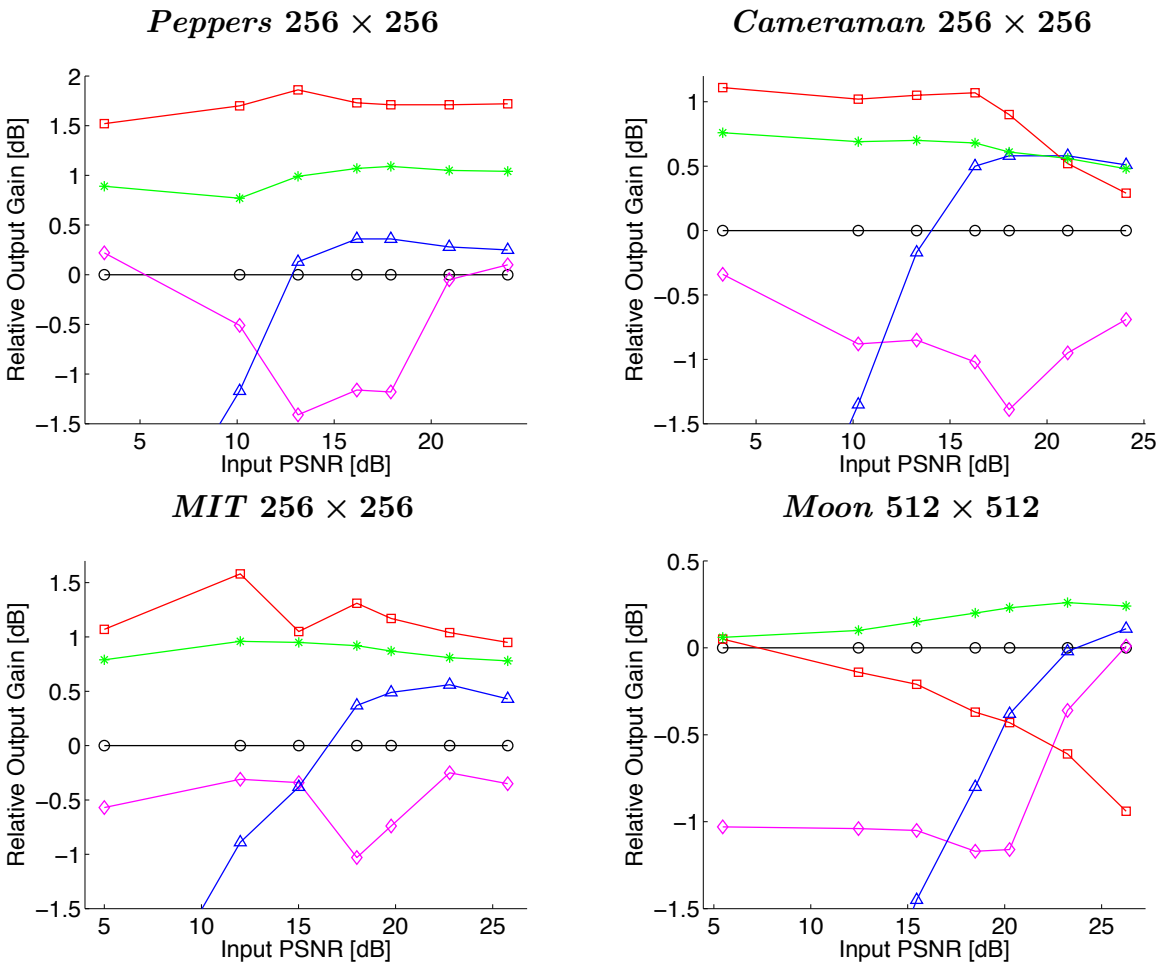


Figure 7.6: Comparison between the non-redundant *PURE-LET*(7.10) ("○" markers), a nearly translation invariant (TI) version of it ("∗" markers), *TI-Haar-Fisz+TI-SUREshrink* ("◊" markers), *GAT+BLS-GSM* applied in a FSP ("△" markers), and *TI-Platelet*("□" markers).

compare our *PURE-LET* algorithm (7.10) with the *GAT+BLS-GSM* applied both in a non-redundant and in a redundant representation. For input PSNRs lower than $\sim$ 12 dB, the VST-based approach is outperformed by the proposed *PURE-LET* estimator, no matter the level of redundancy of the considered transform. For higher PSNRs, the *GAT+BLS-GSM* applied in a full steerable pyramid gives similar results than a nearly shift-invariant (10 CS) version of our *PURE*-based approach.

In Figure 7.10, we show a strongly noisy color image denoised by our multichannel *PURE-LET* algorithm (7.17). As a comparison, we have also displayed the result of the *GAT+BLS-GSM* applied separately on each of the color channel. Beside having

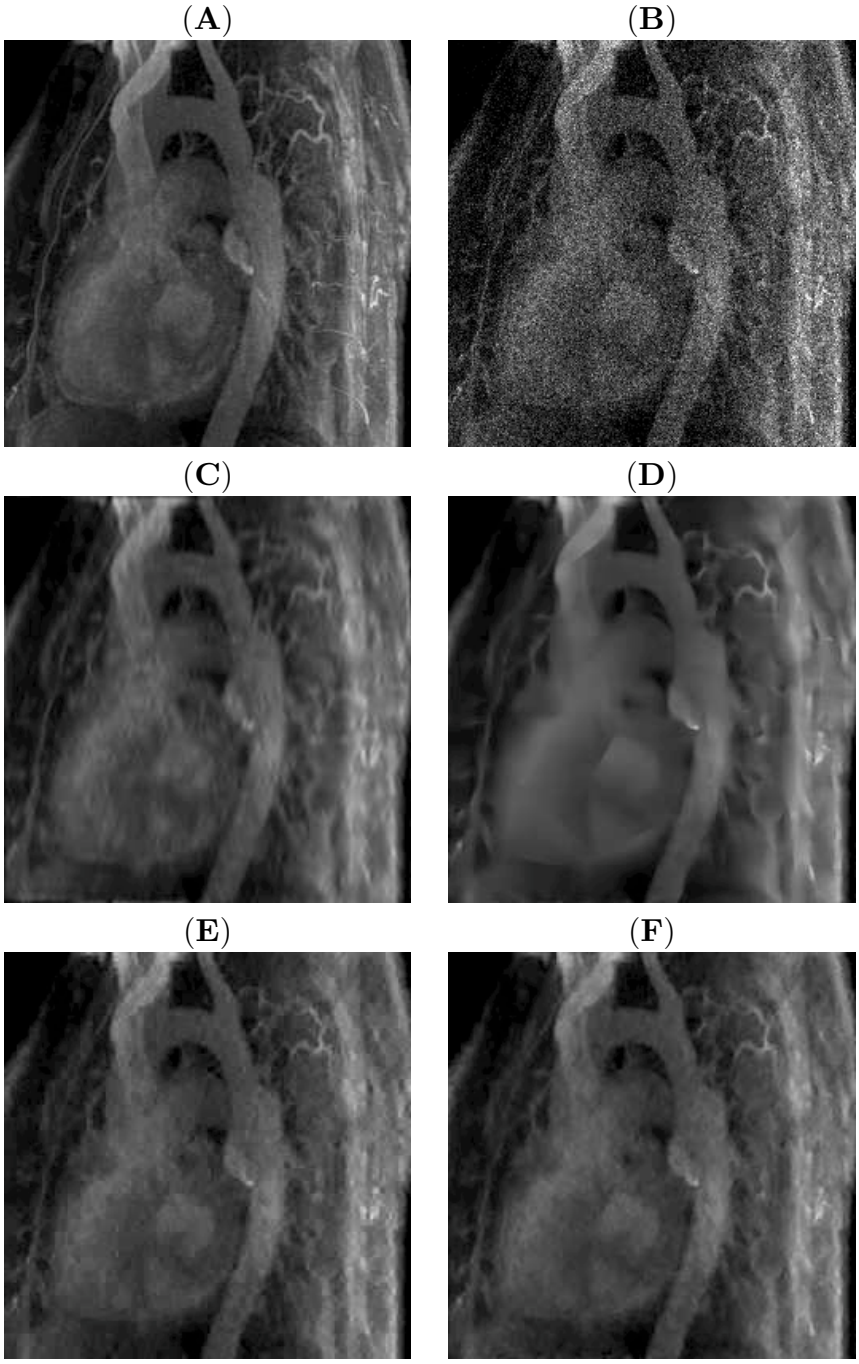


Figure 7.7: (**A**) The original *Pneumatix* image, by courtesy of the following website: `http://pubimage.hcuge.ch:8080/`. (**B**) Noisy realization of it: PSNR = 20.27 dB. (**C**) Denoised with *TI-Haar-Fisz + TI-SUREshrink*: PSNR = 29.08 dB in 1.3s. (**D**) Denoised with *Platelet* (25 CS): PSNR = 29.52 dB in 960s. (**E**) Denoised with our non-redundant *PURE-LET* (7.10): PSNR = 29.21 dB in 0.2s. (**F**) Denoised with our redundant *PURE-LET* (7.10) (25 CS): PSNR = 29.73 dB in 4.4s.

Table 7.2: Cycle-spinning: PSNR improvement vs computation time

***Cameraman* at 17.25 dB**				
Algorithms	**Non-redundant**		**25 cycle-spins**	
	PSNR [dB]	**Time [s]**	**PSNR [dB]**	**Time [s]**
Haar-Fisz + *SUREshrink*	24.01	0.06	24.83	1.3
PURE-LET (7.10)	26.09	0.19	26.75	4.6
Platelet	25.01	44.5	27.17	1112
***MIT* at 18.98 dB**				
Algorithms	**Non-redundant**		**25 cycle-spins**	
	PSNR [dB]	**Time [s]**	**PSNR [dB]**	**Time [s]**
Haar-Fisz + *SUREshrink*	23.47	0.06	24.64	1.3
PURE-LET (7.10)	25.48	0.19	26.43	4.6
Platelet	24.10	36	26.80	891

Note: Output PSNRs and computation times have been averaged over ten noise realizations, except for the *Platelet* approach.

higher PSNRs, we can notice that the *PURE-LET* denoised images are not oversmoothed. For instance, the number "20", which is hardly distinguishable in the noisy image and completely blurred after the *GAT+BLS-GSM* processing, becomes visible after the *PURE-LET* denoising.

7.7 Summary

In this chapter, we have derived an analytical expression of a Haar wavelet domain unbiased MSE estimate for a Poisson process degraded by AWGN (PURE). The unnormalized Haar wavelet transform turns out to be the only multiresolution transform which propagates the Poisson statistics in its lowpass channel.

In the proposed PURE-LET Haar wavelet thresholding, the scaling coefficients play multiple roles. First, they naturally appear in the derivation of PURE, although the latter is only used to optimize the denoising of the *wavelet* coefficients. Second, they serve as an estimate of the signal-dependent variance of the noise contaminating the same-scale wavelet coefficients. Third, they are used to built an interscale predictor for each wavelet coefficients at the same scale.

The performance of the resulting interscale PURE-LET estimator for denoising mono- and multichannel images has been assessed in various simulated experiments. The main conclusions that could be drawn from the comparisons with the state-of-the-art algorithms were the following:

- Although restricted to the maximally decimated Haar wavelet transform, the proposed solution yields results that are comparable or superior to the standard VST-based approaches, even when the latter are applied in a redundant representation.

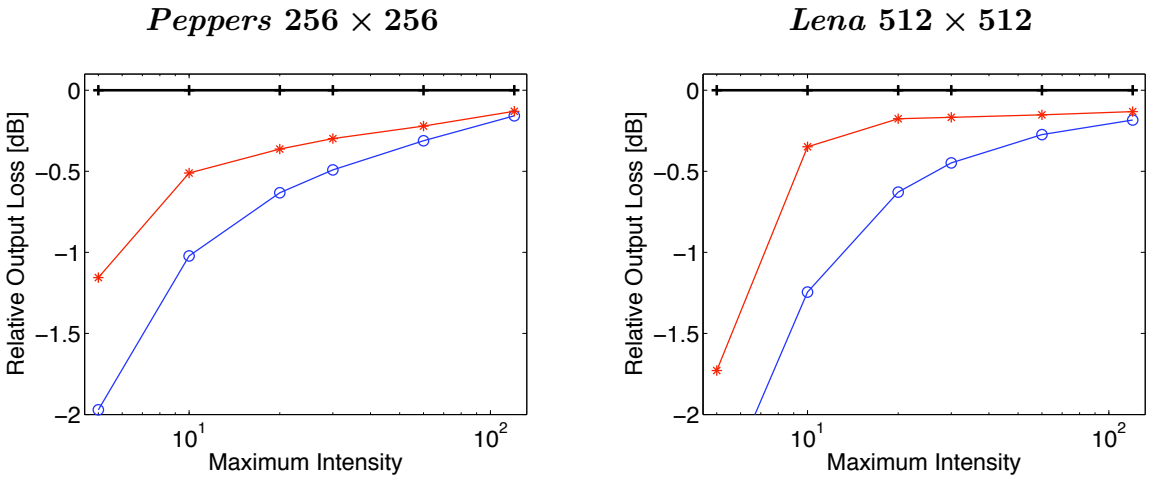


Figure 7.8: PSNR improvements brought by the direct handling of Poisson statistics (*PURE-LET* (OWT Haar): "+" markers) over VST + *SURE-LET* schemes (*Anscombe* + *SURE-LET* (OWT Haar): "∘" markers; *Haar-Fisz* + *SURE-LET* (OWT Haar): "∗" markers).

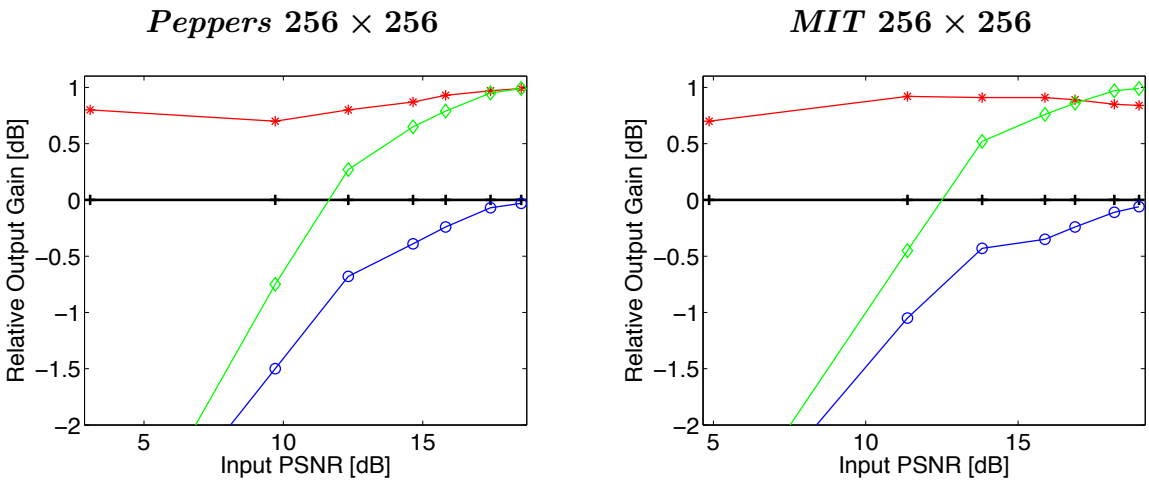


Figure 7.9: Comparison between the non-redundant *PURE-LET*(7.10) ("+" markers), a nearly translation invariant version of it ("∗" markers), and *GAT+BLS-GSM* applied in an orthonormal wavelet representation ("∘" markers) or in a FSP ("⋄" markers).

- A nearly shift-invariant version of our algorithm (obtained by averaging the results of a few cycle-spins) is competitive with the sophisticated *Platelet* method, while being two orders of magnitude faster.

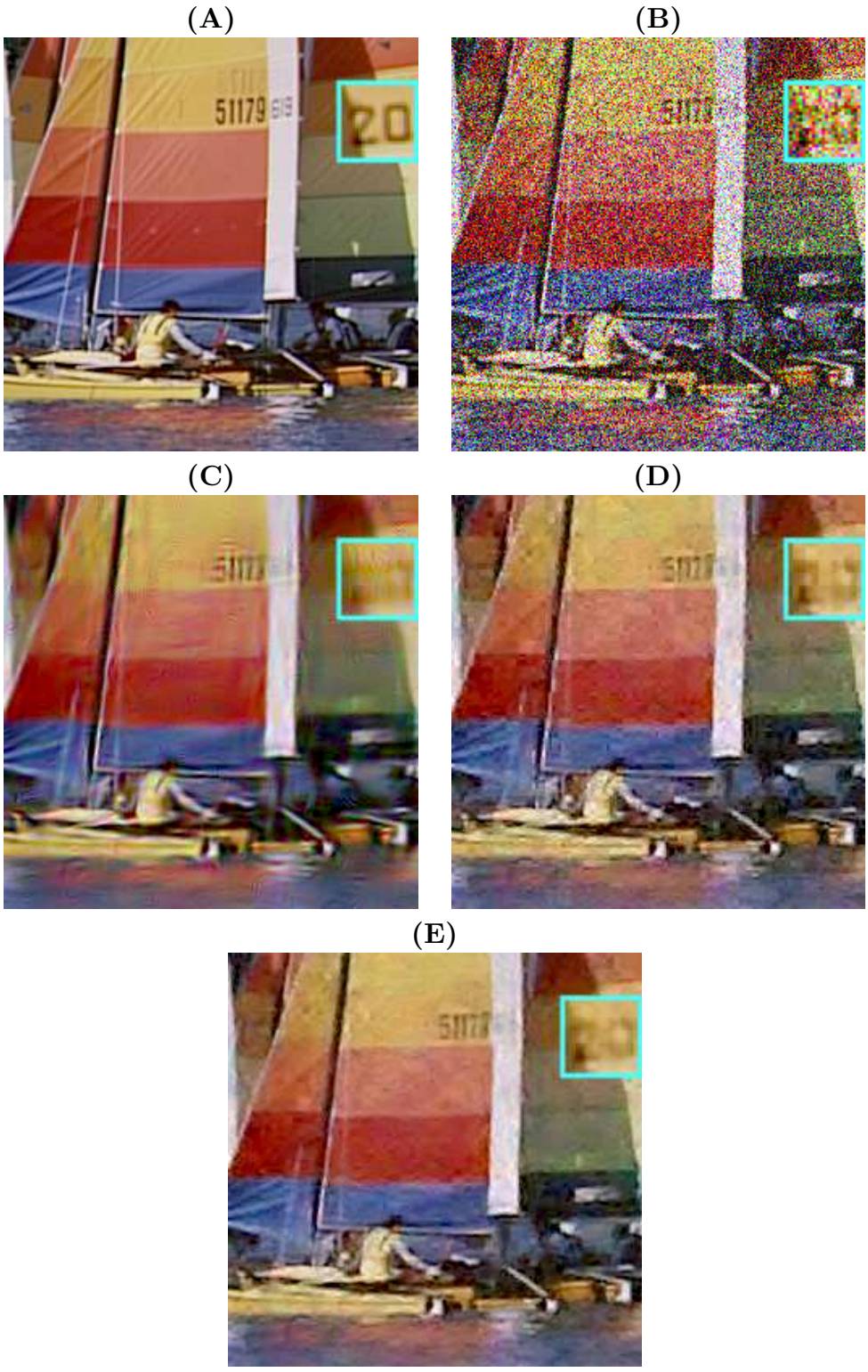


Figure 7.10: (**A**) Part of the original 512×512 *Yacht* image. (**B**) Noisy realization of it: PSNR = 10.77 dB. (**C**) Denoised with *GAT* + *BLS-GSM* (FSP): PSNR = 23.29 dB in 88s. (**D**) Denoised with our non-redundant multichannel *PURE-LET* (7.17): PSNR = 24.35 dB in 4.4s. (**F**) Denoised with a slightly redundant (4 CS) variant of our multichannel *PURE-LET* (7.17): PSNR = 25.23 dB in 17.5s.

Chapter 8

PURE-LET for Arbitrary Transform-Domain Processing

8.1 Motivations

In Chapter 7, we have presented a PURE-LET algorithm restricted to the use of the unnormalized Haar wavelet transform. In this chapter, we propose to generalize the PURE-LET strategy to devise and optimize a broad class of nonlinear processing applied in an *arbitrary* linear transform domain. In particular, our goal is to fully take advantage of the improved denoising quality achievable by the use of redundant transformations.

8.2 Taylor Series Approximation of PURE

PURE, as defined in Theorem 4, is generally time-consuming to evaluate for an arbitrary nonlinear processing, due to the presence of the term $\mathbf{F}^{-}(\mathbf{y})$. Indeed, to compute a single component $f_n(\mathbf{y} - \mathbf{e}_n)$ of $\mathbf{F}^{-}(\mathbf{y})$, one needs to apply the whole denoising process to a slightly perturbed version of the noisy input. This operation has to be repeated N times to get the full vector $\mathbf{F}^{-}(\mathbf{y})$. Such a "brute force" approach is thus prohibitive in practice, considering that a typical image contains $N = 256^2$ pixels. Instead, we propose to use a first order Taylor series expansion of

$\mathbf{F}^-(\mathbf{y})$, i.e. for all $n = 1 \dots N$:

$$f_n(\mathbf{y} - \mathbf{e}_n) = f_n(\mathbf{y}) - \frac{\partial f_n(\mathbf{y})}{\partial y_n} + \underbrace{\int_0^1 (1-t) \frac{\partial^2}{\partial y_n^2} f_n(\mathbf{y} - t\mathbf{e}_n) \mathrm{d}t}_{C_n(\mathbf{y})}$$

where we have assumed that $f_n \in C^2(\mathbb{R}^N)$, $\forall n$.

The remainder $C_n(\mathbf{y})$ of the first order Taylor series approximation of each component of $\mathbf{F}^-(\mathbf{y})$ is upper-bounded by

$$|C_n(\mathbf{y})| \leq \frac{1}{2} \sup_{t \in [0,1]} \left| \frac{\partial^2}{\partial y_n^2} f_n(\mathbf{y} - t\mathbf{e}_n) \right|, \forall n = 1 \dots N \tag{8.1}$$

Consequently, provided that the above upper bounds (8.1) remain negligible compared to the lower order terms, $\mathbf{F}^-(\mathbf{y})$ is well approximated by $\mathbf{F}(\mathbf{y}) - \partial\mathbf{F}(\mathbf{y})$, where $\partial\mathbf{F}(\mathbf{y}) = \mathrm{diag}\left\{\mathbf{J}_{\mathbf{F}}(\mathbf{y})\right\} = [\frac{\partial f_n(\mathbf{y})}{\partial y_n}]_{n=1\dots N}$ is the $N \times 1$ vector made of the first derivative of each function f_n with respect to y_n. The PURE unbiased MSE estimate defined in (6.3) is, in turn, well approximated by:

$$\begin{aligned} \tilde{\epsilon} = & \frac{1}{N} \left(\|\mathbf{F}(\mathbf{y})\|^2 - 2\mathbf{y}^{\mathrm{T}}(\mathbf{F}(\mathbf{y}) - \partial\mathbf{F}(\mathbf{y})) + 2\sigma^2 \mathrm{div}\left\{\mathbf{F}(\mathbf{y}) - \partial\mathbf{F}(\mathbf{y})\right\} \right) + \\ & \frac{1}{N}(\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}) - \sigma^2 \end{aligned} \tag{8.2}$$

If the processing $\mathbf{F}$ is *linear*, then the two MSE estimates (6.3) and (8.2) are obviously equivalent.

Note that a first order Taylor series approximation was also proposed by Yuan in [170] to derive a computable unbiased estimate of the Kullback-Leibler distance[1] for a generalized Poisson process. The so-built unbiased estimate was then used to optimize the smoothing parameter of a penalized likelihood estimator of the Poisson intensities.

8.3 PURE for Transform-Domain Denoising

In this section, we derive an expression for the approximated PURE given in Equ. (8.2), in the particular case of a transformed domain processing $\mathbf{\Theta}$. Dealing with signal-dependent noise makes it generally difficult to express the observation model in the transformed domain. Therefore, we assume here that a (coarse) "map" of the signal-dependent noise variance can be obtained in the transformed domain, by applying a linear transformation (typically a smoothing) $\widetilde{\mathbf{D}} = [\tilde{d}_{i,j}]_{(i,j) \in [1;L] \times [1;N]}$ to the noisy

[1] The Kullback-Leibler (KL) distance is a non-symmetric measure between two continuous (resp. discrete) PDF p_1 and p_2, defined as: $\mathrm{KL}(p_1, p_2) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)} \mathrm{d}x$ (resp. $\sum_i p_1(x_i) \log \frac{p_1(x_i)}{p_2(x_i)}$).

data $\mathbf{y}$. The denoised estimate $\hat{\mathbf{x}}$ can be thus expressed as a function $\mathbf{F}$ of the noisy signal $\mathbf{y}$:

$$\hat{\mathbf{x}} = \mathbf{F}(\mathbf{y}) = \mathbf{R}\boldsymbol{\Theta}(\underbrace{\mathbf{D}\mathbf{y}}_{\mathbf{w}}, \underbrace{\widetilde{\mathbf{D}}\mathbf{y}}_{\widetilde{\mathbf{w}}}) \tag{8.3}$$

where $\boldsymbol{\Theta}(\mathbf{w}, \widetilde{\mathbf{w}}) = [\theta_l(\mathbf{w}, \widetilde{\mathbf{w}})]_{l\in[1;L]}$ is a possibly multivariate processing.

Note that the *PUREshrink* and the PURE-LET algorithms introduced in Sections 7.4 and 7.5 belong to the general class of processing defined in Equ. (8.3). In their case, $\mathbf{D}$ (resp. $\mathbf{R}$) implements the (unnormalized) Haar wavelet decomposition (resp. reconstruction) and $\mathbf{w}$ (resp. $\widetilde{\mathbf{w}}$) represents the Haar wavelet (resp. scaling) coefficients.

Corollary 8. *For a transform-domain processing as defined in* (8.3)*, the first order approximation of the MSE estimate given in Equ.* (8.2) *can be reformulated as:*

$$\begin{aligned}\tilde{\epsilon} \;=\; & \frac{1}{N}\left(\|\mathbf{F}(\mathbf{y})-\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right) - \sigma^2 + \\ & \frac{2}{N}\mathbf{y}^{\mathrm{T}}\mathrm{diag}\left\{\mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\mathbf{w})\mathbf{D} + \mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\widetilde{\mathbf{w}})\widetilde{\mathbf{D}}\right\} + \\ & \frac{2\sigma^2}{N}\mathrm{trace}\left\{\mathbf{D}\mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\mathbf{w}) + \widetilde{\mathbf{D}}\mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\widetilde{\mathbf{w}})\right\} - \\ & \frac{2\sigma^2}{N}\mathrm{div}\left\{\mathrm{diag}\left\{\mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\mathbf{w})\mathbf{D} + \mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\widetilde{\mathbf{w}})\widetilde{\mathbf{D}}\right\}\right\}\end{aligned} \tag{8.4}$$

where

- $\mathbf{J}_{\boldsymbol{\Theta}}(\mathbf{w}) = \left[\frac{\partial\theta_k(\mathbf{w},\widetilde{\mathbf{w}})}{\partial w_l}\right]_{1\le k,l\le L}$ *is the* $L \times L$ *Jacobian matrix of the transform-domain processing* $\boldsymbol{\Theta}$*, made of the first-order partial derivative of each function* $\theta_k(\mathbf{w}, \widetilde{\mathbf{w}})$ *w.r.t. the components of its first variable* $\mathbf{w}$*.*

- $\mathbf{J}_{\boldsymbol{\Theta}}(\widetilde{\mathbf{w}}) = \left[\frac{\partial\theta_k(\mathbf{w},\widetilde{\mathbf{w}})}{\partial \widetilde{w}_l}\right]_{1\le k,l\le L}$ *is the* $L \times L$ *Jacobian matrix of the transform-domain processing* $\boldsymbol{\Theta}$*, made of the first-order partial derivative of each function* $\theta_k(\mathbf{w}, \widetilde{\mathbf{w}})$ *w.r.t. the components of its second variable* $\widetilde{\mathbf{w}}$*.*

Proof. Using the result given in (8.2), we only have to further develop the Jacobian matrix of the denoising process $\mathbf{J}_{\mathbf{F}}(\mathbf{y}) = [\frac{\partial f_i(\mathbf{y})}{\partial y_j}]_{1\le i,j\le N}$ which appears in $\mathbf{y}^{\mathrm{T}}\partial\mathbf{F}(\mathbf{y}) = \mathbf{y}^{\mathrm{T}}\mathrm{diag}\{\mathbf{J}_{\mathbf{F}}(\mathbf{y})\}$, $\mathrm{div}\{\mathbf{F}(\mathbf{y})\} = \mathrm{trace}\{\mathbf{J}_{\mathbf{F}}(\mathbf{y})\}$ and $\mathrm{div}\{\partial\mathbf{F}(\mathbf{y})\} =$

$\mathrm{div}\,\{\mathrm{diag}\,\{\mathbf{J_F}(\mathbf{y})\}\}$:

$$\begin{aligned}
\frac{\partial f_i(\mathbf{y})}{\partial y_j} &= \sum_{l=1}^{L} r_{i,l} \frac{\partial \theta_l(\mathbf{w}, \widetilde{\mathbf{w}})}{\partial y_j} \\
&= \sum_{l=1}^{L} r_{i,l} \sum_{k=1}^{L} \Big(\frac{\partial \theta_l(\mathbf{w}, \widetilde{\mathbf{w}})}{\partial w_k} \underbrace{\frac{\partial w_k}{\partial y_j}}_{d_{k,j}} + \frac{\partial \theta_l(\mathbf{w}, \widetilde{\mathbf{w}})}{\partial \widetilde{w}_k} \underbrace{\frac{\partial \widetilde{w}_k}{\partial y_j}}_{\tilde{d}_{k,j}} \Big) \\
&= \sum_{l=1}^{L} \sum_{k=1}^{L} r_{i,l} \Big(\underbrace{\frac{\partial \theta_l(\mathbf{w}, \widetilde{\mathbf{w}})}{\partial w_k}}_{[\mathbf{J}_{\boldsymbol{\Theta}}(\mathbf{w})]_{l,k}} d_{k,j} + \underbrace{\frac{\partial \theta_l(\mathbf{w}, \widetilde{\mathbf{w}})}{\partial \widetilde{w}_k}}_{[\mathbf{J}_{\boldsymbol{\Theta}}(\widetilde{\mathbf{w}})]_{l,k}} \tilde{d}_{k,j} \Big) \qquad (8.5)
\end{aligned}$$

Hence, $\mathbf{J_F}(\mathbf{y}) = \mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\mathbf{w})\mathbf{D} + \mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\widetilde{\mathbf{w}})\widetilde{\mathbf{D}}$ and consequently,

$$\begin{aligned}
\mathbf{y}^{\mathrm{T}} \partial \mathbf{F}(\mathbf{y}) &= \mathbf{y}^{\mathrm{T}} \mathrm{diag} \Big\{ \mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\mathbf{w})\mathbf{D} + \mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\widetilde{\mathbf{w}})\widetilde{\mathbf{D}} \Big\} \qquad (8.6) \\
\mathrm{div}\,\{\mathbf{F}(\mathbf{y})\} &= \mathrm{trace} \Big\{ \mathbf{D}\mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\mathbf{w}) + \widetilde{\mathbf{D}}\mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\widetilde{\mathbf{w}}) \Big\} \qquad (8.7) \\
\mathrm{div}\,\{\partial \mathbf{F}(\mathbf{y})\} &= \mathrm{div} \Big\{ \mathrm{diag} \Big\{ \mathbf{D}\mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\mathbf{w}) + \widetilde{\mathbf{D}}\mathbf{R}\mathbf{J}_{\boldsymbol{\Theta}}(\widetilde{\mathbf{w}}) \Big\} \Big\} \qquad (8.8)
\end{aligned}$$

Putting back Equ. (8.6), (8.7) and (8.8) into formula (8.2) finally gives the desired result (8.4). ■

8.4 Pointwise Estimator for Undecimated Filterbank Transforms

In this section, we show how to use the general result of Corollary 8 to globally optimize a simple *pointwise* thresholding applied to the coefficients of a J-band undecimated filterbank (similar to the one depicted in Figure 4.1 and described in Section 4.3.2). Although all the results presented in this section are derived for 1D signals, they can be straightforwardly extended to higher dimensional signals, when considering separable filters.

8.4.1 Pointwise PURE

In Corollary 8, we gave the expression of the first order Taylor series approximation of PURE for an arbitrary transformed domain processing $\boldsymbol{\Theta}$. We now restrict our study to the particular case of a *pointwise* processing, i.e. $\boldsymbol{\Theta}(\mathbf{w}, \widetilde{\mathbf{w}}) = [\theta_l(w_l, \widetilde{w}_l)]_{l \in [1;L]}$.

Corollary 9. *For a transformed domain* pointwise *processing, PURE can be approximated by:*

$$\begin{aligned}\tilde{\epsilon} \;=\; & \frac{1}{N}\left(\|\mathbf{F}(\mathbf{y})-\mathbf{y}\|^2-\mathbf{1}^{\mathrm{T}}\mathbf{y}\right)-\sigma^2+\\ & \frac{2}{N}\Big(\boldsymbol{\Theta}_1(\mathbf{w},\widetilde{\mathbf{w}})^{\mathrm{T}}(\mathbf{D}\bullet\mathbf{R}^{\mathrm{T}})\mathbf{y}+\boldsymbol{\Theta}_2(\mathbf{w},\widetilde{\mathbf{w}})^{\mathrm{T}}(\widetilde{\mathbf{D}}\bullet\mathbf{R}^{\mathrm{T}})\mathbf{y}\Big)+\\ & \frac{2\sigma^2}{N}\Big(\mathrm{diag}\left\{\mathbf{DR}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_1(\mathbf{w},\widetilde{\mathbf{w}})+\mathrm{diag}\left\{\widetilde{\mathbf{D}}\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_2(\mathbf{w},\widetilde{\mathbf{w}})\\ & -\mathrm{diag}\left\{\mathbf{D}^2\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_{11}(\mathbf{w},\widetilde{\mathbf{w}})-\mathrm{diag}\left\{\widetilde{\mathbf{D}}^2\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_{22}(\mathbf{w},\widetilde{\mathbf{w}})\\ & -2\,\mathrm{diag}\left\{(\mathbf{D}\bullet\widetilde{\mathbf{D}})\mathbf{R}\right\}^{\mathrm{T}}\boldsymbol{\Theta}_{12}(\mathbf{w},\widetilde{\mathbf{w}})\Big)\end{aligned} \tag{8.9}$$

where

- $\boldsymbol{\Theta}_1(\mathbf{w},\widetilde{\mathbf{w}})=\left[\frac{\partial\theta_l(w_l,\widetilde{w}_l)}{\partial w_l}\right]_{l\in[1;L]}$ *is the* $L\times 1$ *vector made of the first derivative, with respect to its first variable, of each function* θ_l*.*
- $\boldsymbol{\Theta}_2(\mathbf{w},\widetilde{\mathbf{w}})=\left[\frac{\partial\theta_l(w_l,\widetilde{w}_l)}{\partial \widetilde{w}_l}\right]_{l\in[1;L]}$ *is the* $L\times 1$ *vector made of the first derivative, with respect to its second variable, of each function* θ_l*.*
- $\boldsymbol{\Theta}_{12}(\mathbf{w},\widetilde{\mathbf{w}})=\left[\frac{\partial^2\theta_l(w_l,\widetilde{w}_l)}{\partial w_l\partial \widetilde{w}_l}\right]_{l\in[1;L]}$ *is the* $L\times 1$ *vector made of the first derivative, with respect to its first variable and second variable, of each function* θ_l*.*
- $\boldsymbol{\Theta}_{11}(\mathbf{w},\widetilde{\mathbf{w}})=\left[\frac{\partial^2\theta_l(w_l,\widetilde{w}_l)}{\partial w_l^2}\right]_{l\in[1;L]}$ *is the* $L\times 1$ *vector made of the second derivative, with respect to its first variable, of each function* θ_l*.*
- $\boldsymbol{\Theta}_{22}(\mathbf{w},\widetilde{\mathbf{w}})=\left[\frac{\partial^2\theta_l(w_l,\widetilde{w}_l)}{\partial \widetilde{w}_l^2}\right]_{l\in[1;L]}$ *is the* $L\times 1$ *vector made of the second derivative, with respect to its second variable, of each function* θ_l*.*
- *"*$\bullet$*" denotes the Hadamard (element-by-element) product between two matrices.*
- $\mathbf{D}^2$ *(resp.* $\widetilde{\mathbf{D}}^2$*) stands for* $\mathbf{D}\bullet\mathbf{D}$ *(resp.* $\widetilde{\mathbf{D}}\bullet\widetilde{\mathbf{D}}$*).*

Proof. Setting $i=j$, $k=l$ and replacing $\theta_l(\mathbf{w},\widetilde{\mathbf{w}})$ by its pointwise restriction $\theta_l(w_l,\widetilde{w}_l)$ in Equ. (8.5) leads to:

$$\frac{\partial f_i(\mathbf{y})}{\partial y_i}=\sum_{l=1}^{L} r_{i,l}\Big(\underbrace{\frac{\partial\theta_l(w_l,\widetilde{w}_l)}{\partial w_l}}_{[\boldsymbol{\Theta}_1(\mathbf{w},\widetilde{\mathbf{w}})]_l}d_{l,i}+\underbrace{\frac{\partial\theta_l(w_l,\widetilde{w}_l)}{\partial \widetilde{w}_l}}_{[\boldsymbol{\Theta}_2(\mathbf{w},\widetilde{\mathbf{w}})]_l}\tilde{d}_{l,i}\Big)$$

Similarly, we have:

$$\frac{\partial^2 f_i(\mathbf{y})}{\partial y_i^2} = \sum_{l=1}^{L} r_{i,l} \Big(\underbrace{\frac{\partial^2 \theta_l(w_l, \widetilde{w}_l)}{\partial w_l^2}}_{[\mathbf{\Theta}_{11}(\mathbf{w},\widetilde{\mathbf{w}})]_l} d_{l,i}^2 + \underbrace{\frac{\partial^2 \theta_l(w_l, \widetilde{w}_l)}{\partial \widetilde{w}_l^2}}_{[\mathbf{\Theta}_{22}(\mathbf{w},\widetilde{\mathbf{w}})]_l} \tilde{d}_{l,i}^2 + 2 \underbrace{\frac{\partial^2 \theta_l(w_l, \widetilde{w}_l)}{\partial w_l \partial \widetilde{w}_l}}_{[\mathbf{\Theta}_{12}(\mathbf{w},\widetilde{\mathbf{w}})]_l} d_{l,i} \tilde{d}_{l,i} \Big) \tag{8.10}$$

and consequently:

$$\begin{aligned} \mathbf{y}^{\mathrm{T}} \partial \mathbf{F}(\mathbf{y}) &= \sum_{l=1}^{L} [\mathbf{\Theta}_1(\mathbf{w}, \widetilde{\mathbf{w}})]_l \sum_{i=1}^{N} \underbrace{d_{l,i} r_{i,l}}_{[\mathbf{D} \bullet \mathbf{R}^{\mathrm{T}}]_{l,i}} y_i + \sum_{l=1}^{L} [\mathbf{\Theta}_2(\mathbf{w}, \widetilde{\mathbf{w}})]_l \sum_{i=1}^{N} \underbrace{\tilde{d}_{l,i} r_{i,l}}_{[\widetilde{\mathbf{D}} \bullet \mathbf{R}^{\mathrm{T}}]_{l,i}} y_i \\ &= \mathbf{\Theta}_1(\mathbf{w}, \widetilde{\mathbf{w}})^{\mathrm{T}} (\mathbf{D} \bullet \mathbf{R}^{\mathrm{T}}) \mathbf{y} + \mathbf{\Theta}_2(\mathbf{w}, \widetilde{\mathbf{w}})^{\mathrm{T}} (\widetilde{\mathbf{D}} \bullet \mathbf{R}^{\mathrm{T}}) \mathbf{y} \end{aligned} \tag{8.11}$$

$$\begin{aligned} \operatorname{div}\{\mathbf{F}(\mathbf{y})\} &= \sum_{l=1}^{L} [\mathbf{\Theta}_1(\mathbf{w}, \widetilde{\mathbf{w}})]_l \sum_{i=1}^{N} \underbrace{d_{l,i} r_{i,l}}_{[\mathbf{DR}]_{l,l}} + \sum_{l=1}^{L} [\mathbf{\Theta}_2(\mathbf{w}, \widetilde{\mathbf{w}})]_l \sum_{i=1}^{N} \underbrace{\tilde{d}_{l,i} r_{i,l}}_{[\widetilde{\mathbf{D}}\mathbf{R}]_{l,l}} \\ &= \operatorname{diag}\{\mathbf{DR}\}^{\mathrm{T}} \mathbf{\Theta}_1(\mathbf{w}, \widetilde{\mathbf{w}}) + \operatorname{diag}\left\{\widetilde{\mathbf{D}}\mathbf{R}\right\}^{\mathrm{T}} \mathbf{\Theta}_2(\mathbf{w}, \widetilde{\mathbf{w}}) \end{aligned} \tag{8.12}$$

$$\begin{aligned} \operatorname{div}\{\partial \mathbf{F}(\mathbf{y})\} &= \sum_{l=1}^{L} [\mathbf{\Theta}_{11}(\mathbf{w}, \widetilde{\mathbf{w}})]_l \sum_{i=1}^{N} \underbrace{d_{l,i}^2 r_{i,l}}_{[\mathbf{D}^2\mathbf{R}]_{l,l}} + \sum_{l=1}^{L} [\mathbf{\Theta}_{22}(\mathbf{w}, \widetilde{\mathbf{w}})]_l \sum_{i=1}^{N} \underbrace{\tilde{d}_{l,i}^2 r_{i,l}}_{[\widetilde{\mathbf{D}}^2\mathbf{R}]_{l,l}} + \\ &\quad 2 \sum_{l=1}^{L} [\mathbf{\Theta}_{12}(\mathbf{w}, \widetilde{\mathbf{w}})]_l \sum_{i=1}^{N} \underbrace{d_{l,i} \tilde{d}_{l,i} r_{i,l}}_{[(\mathbf{D} \bullet \widetilde{\mathbf{D}})\mathbf{R}]_{l,l}} \\ &= \operatorname{diag}\left\{\mathbf{D}^2\mathbf{R}\right\}^{\mathrm{T}} \mathbf{\Theta}_{11}(\mathbf{w}, \widetilde{\mathbf{w}}) + \operatorname{diag}\left\{\widetilde{\mathbf{D}}^2\mathbf{R}\right\}^{\mathrm{T}} \mathbf{\Theta}_{22}(\mathbf{w}, \widetilde{\mathbf{w}}) + \\ &\quad 2 \operatorname{diag}\left\{(\mathbf{D} \bullet \widetilde{\mathbf{D}})\mathbf{R}\right\}^{\mathrm{T}} \mathbf{\Theta}_{12}(\mathbf{w}, \widetilde{\mathbf{w}}) \end{aligned} \tag{8.13}$$

Putting back Equ. (8.11), (8.12) and (8.13) into formula (8.2) finally gives the desired result (8.9). ■

8.4.2 Choice of $\widetilde{\mathbf{D}}$

In a multiscale undecimated filterbank, the scaling coefficients of the lowpass residual at a given scale j could be used as a reliable estimation of the signal-dependent noise variance[2] for each bandpass subband from the same scale j. However, except for

[2] Up to a scale-dependent factor β_j given by: $\beta_j = \dfrac{\sum_k \tilde{g}_j[k]^2}{\sum_k \tilde{h}_j[k]}$.

the Haar wavelet filters, there is a group delay difference between the output of the lowpass and bandpass channels that needs to be compensated for. Contrary to the construction of the interscale predictor described in Section 3.3.1, we do not want the output of the lowpass channel to "look like" the output of the bandpass channel. We are therefore looking for a *lowpass* filter $Q(z)$, such that:

$$H(z^{-1})Q(z) = G(-z^{-1})R_1(z), \text{ where } R_1(z) = \pm R_1(z^{-1}) \text{ is arbitrary.}$$

Under these design constraints, and using the fact that $G(-z^{-1}) = zH(z)$, one can show that the desired group delay compensation (GDC) filter is given by:

$$Q(z) = zH(z)^2R(z), \text{ where } R(z) = \pm R(z^{-1}) \text{ is arbitrary.} \tag{8.14}$$

In practice, we will consider a normalized filter $Q(z)$, i.e. such that $\sum_{n\in\mathbb{Z}} q_n^2 = 1$. Note that, in the case of symmetric or nearly symmetric filters, such as the Daubechies *symlets*, the shortest-support GDC filter is simply given by $Q(z) = z$, i.e. a one-sample shift.

The $L \times N$ matrix $\widetilde{\mathbf{D}}$ can be then defined as $\widetilde{\mathbf{D}} = [\widetilde{\mathbf{D}}_1^{\mathrm{T}}\ \widetilde{\mathbf{D}}_2^{\mathrm{T}} \dots \widetilde{\mathbf{D}}_{J+1}^{\mathrm{T}}]^{\mathrm{T}}$, with

$$\left[\widetilde{\mathbf{D}}_j\right]_{k,l} = \sum_{n\in\mathbb{Z}} \tilde{h}_j[l-k+nN] \ ,$$

where $\tilde{h}_j[n]$ is the n^{th} coefficient of the filter

$$\begin{cases} \tilde{H}_j(z) &= 2^j Q(z^{2^{j-1}})H_j(z) = Q(z^{2^{j-1}})H(z)H(z^2)\dots H(z^{2^{j-1}}), \text{ for } j = 1, 2, \dots, J. \\ \tilde{H}_{J+1}(z) &= 2^J H_J(z) = H(z)H(z^2)\dots H(z^{2^{J-1}}) \end{cases}$$

In an overcomplete BDCT representation, the lowpass residual band can directly serve as a coarse estimate of the noise variance[3] for each bandpass subband, since the filters of the BDCT have the same group delay.

8.4.3 Computation of Transform-dependent Terms

To compute the approximated unbiased MSE estimate of Equ. (8.9), we need to evaluate several terms that solely depend on the choice of transformation:

1. Computation of diag $\{\mathbf{DR}\}$:
   We have already shown in Section 4.3.2 that, thanks to the biorthogonality condition (4.9),

$$\text{diag}\,\{\mathbf{D}_j\mathbf{R}_j\} = \begin{cases} \dfrac{1}{M_j}[\underbrace{1, 1, \dots, 1}_{N \text{ times}}]^{\mathrm{T}} & \text{, for } j = 1, 2, \dots, J \\ \dfrac{1}{M_J}[\underbrace{1, 1, \dots, 1}_{N \text{ times}}]^{\mathrm{T}} & \text{, for } j = J+1 \end{cases}$$

[3] Up to the scaling factor $\beta_j = 1/\sqrt{M}$.

where for multiscale filterbanks, $M_j = 2^j$ is the downsampling factor. For an overcomplete BDCT, $M_j = M$, where M is the size of the considered blocks.

2. Computation of diag $\left\{\widetilde{\mathbf{D}}\mathbf{R}\right\}$, diag $\left\{\mathbf{D}^2\mathbf{R}\right\}$, diag $\left\{\widetilde{\mathbf{D}}^2\mathbf{R}\right\}$ and diag $\left\{(\mathbf{D} \bullet \widetilde{\mathbf{D}})\mathbf{R}\right\}$: Contrary to the previous diagonal term, the computation of the above diagonal terms does not generally lead to simple expressions that are independent from the coefficients of the underlying filters. However, all diagonal terms presented in this section can be easily computed offline, using the numerical algorithm described in Section 4.3.4. In particular, when the various matrices are not explicitly given or when non-periodic boundary extensions are performed, this numerical scheme can also be applied to compute all these diagonal terms.

Note that the vector $(\mathbf{D} \bullet \mathbf{R}^{\mathrm{T}})\mathbf{y}$ (resp. $(\widetilde{\mathbf{D}} \bullet \mathbf{R}^{\mathrm{T}})\mathbf{y}$) is obtained by analyzing the noisy data $\mathbf{y}$ with the considered undecimated filterbank using modified analysis filters with coefficients $\tilde{g}_j[k]g_j[-k]$ (resp. $\tilde{h}_j[k]g_j[-k]$).

8.4.4 Thresholding Function

In the case of a Poisson process, we need a signal-dependent transform-domain threshold to take into account the non-stationarity of the noise. If we consider unit-norm filters, the variance σ^2 of the AWGN is preserved in the transformed domain. An estimation of the variance of the Poisson noise component is given by the magnitude of $|\widetilde{w}_j|$ (built as described in Section 8.4.2), up to the scale-dependent factor $\beta_j = 2^{-j/2}$ (resp. $\beta_j = 1/\sqrt{M}$) for a multiscale transform (resp. for an overcomplete BDCT). Therefore, we propose the following signal-dependent threshold, $t_j(\widetilde{w}) = \sqrt{\beta_j|\widetilde{w}| + \sigma^2}$, which is then embedded in a subband-dependent thresholding function, similarly to the one proposed in Section 4.3.3 for AWGN reduction in redundant representations:

$$\theta_j(w, \widetilde{w}) = a_{j,1} \underbrace{w}_{\theta_{j,1}(w,\widetilde{w})} + a_{j,2} \underbrace{we^{-\left(\frac{w}{3t_j(\widetilde{w})}\right)^8}}_{\theta_{j,2}(w,\widetilde{w})} \tag{8.15}$$

To compute the approximated PURE given in Equ. (8.9), one needs to use a differentiable (at least, up to the second order) approximation of the absolute value function: in practice, we suggest to use $\tanh(k\widetilde{w})\widetilde{w}$ (typically $k = 100$) instead of $|\widetilde{w}|$.

As observed in Figure 8.1, the proposed thresholding function can be perceived as a smooth hard-thresholding. Thanks to the use of a signal-dependent threshold, each transformed coefficient is adaptively thresholded according to its estimated amount of noise.

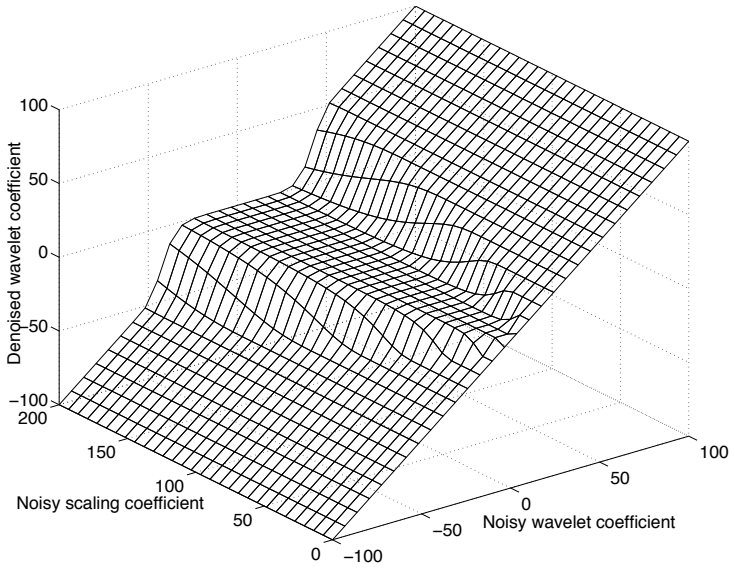


Figure 8.1: A possible realization of the proposed thresholding function (8.15) in a particular subband of an undecimated Haar wavelet representation.

8.4.5 Reliability

In the generalized PURE-LET framework, the whole transformed domain thresholding can be rewritten as:

$$\mathbf{F}(\mathbf{y}) = \sum_{j=1}^{J}\sum_{k=1}^{2} a_{j,k} \underbrace{\mathbf{R}_j \boldsymbol{\Theta}_{j,k}(\mathbf{D}_j\mathbf{y}, \widetilde{\mathbf{D}}_j\mathbf{y})}_{\mathbf{F}_{j,k}(\mathbf{y})} + \underbrace{\mathbf{R}_{J+1}\mathbf{D}_{J+1}\mathbf{y}}_{\text{lowpass}} \tag{8.16}$$

The parameters $a_{j,k}$ that minimize the approximate MSE estimate $\tilde{\epsilon}$ given in Equ. (8.2) are then the solution of the linear system of $2J$ equations $\mathbf{Ma} = \tilde{\mathbf{c}}$, where for $\mathbf{k}, \mathbf{l} \in [1; J] \times [1; 2]$,

$$\begin{cases} \mathbf{M} &= \left[\mathbf{F_k}(\mathbf{y})^{\mathrm{T}}\mathbf{F_l}(\mathbf{y})\right]_{2(k_1-1)+k_2,2(l_1-1)+l_2} \\ \tilde{\mathbf{c}} &= \Big[\mathbf{F_k}(\mathbf{y})^{\mathrm{T}}(\mathbf{Id} - \mathbf{R}_{J+1}\mathbf{D}_{J+1})\mathbf{y} - \partial\mathbf{F_k}(\mathbf{y})^{\mathrm{T}}\mathbf{y} \\ & \quad -\sigma^2 \mathrm{div}\left\{\mathbf{F_k}(\mathbf{y}) - \partial\mathbf{F_k}(\mathbf{y})\right\}\Big]_{2(k_1-1)+k_2} \end{cases} \tag{8.17}$$

For very low intensity signals (typically such that $x_n < 5$, $\forall n$), the first order Taylor series approximation of some nonlinear functions $\mathbf{F}_{j,2}^{-}$ might be inaccurate,

leading to an unreliable approximation of the $(j,2)^{\text{th}}$ component of the PURE vector $\mathbf{c}$. To illustrate this scenario (in the case of an undecimated Haar wavelet representation), we have plotted in the first row of Figure 8.2 the SNR between $c_{j,2}$ (for $j=1$ and $j=3$) and its approximation $\tilde{c}_{j,2}$ as a function of the average noise-free signal energy $E_{\text{mean}} = 1/N \; \|\mathbf{x}\|^2 = 1/N \; \mathscr{E}\left\{\|\mathbf{y}\|^2 - \mathbf{1}^{\mathrm{T}}\mathbf{y}\right\} - \sigma^2$. Any non-linearly processed subband $\mathbf{F}_{j,2}(\mathbf{y})$ for which the SNR of the approximation is below a given threshold (typically around 40 dB), should be disregarded from the LET. In practice, we need to identify the badly-approximated functions without computing their corresponding SNR. Hereafter, we detail two approaches for this identification.

The first consists in monitoring the decrease of the approximation quality by the value of the upper-bound on the Taylor series remainder proposed in Equ. (8.1). In the particular case of a transformed-domain *pointwise* processing, this upper-bound $|C_n(\mathbf{y})|$ can be further developed as:

$$
\begin{aligned}
&\sup_{t\in[0,1]} \left| \frac{\partial^2}{\partial y_n^2} f_n(\mathbf{y} - t\mathbf{e}_n) \right| \\
&\overset{(8.10)}{\leq} \sup_{t\in[0,1]} \sum_{l=1}^{L} \left| r_{n,l} \left(d_{l,n}^2 \frac{\partial^2}{\partial w_l^2} \theta_l(w_l - t d_{l,n}, \widetilde{w}_l - t\tilde{d}_{l,n}) + \tilde{d}_{l,n}^2 \frac{\partial^2}{\partial \widetilde{w}_l^2} \theta_l(w_l - t d_{l,n}, \widetilde{w}_l - t\tilde{d}_{l,n}) \right.\right. \\
&\qquad \left.\left. +2\, d_{l,n}\tilde{d}_{l,n} \frac{\partial^2}{\partial w_l \partial \widetilde{w}_l} \theta_l(w_l - t d_{l,n}, \widetilde{w}_l - t\tilde{d}_{l,n}) \right) \right| \\
&\leq \sup_{t\in[0,1]} \sum_{l=1}^{L} \left(\left| \frac{\partial^2}{\partial w_l^2} \theta_l(w_l - t d_{l,n}, \widetilde{w}_l - t\tilde{d}_{l,n}) \right| + \left| \frac{\partial^2}{\partial \widetilde{w}_l^2} \theta_l(w_l - t d_{l,n}, \widetilde{w}_l - t\tilde{d}_{l,n}) \right| \right. \\
&\qquad \left. +2 \left| \frac{\partial^2}{\partial w_l \partial \widetilde{w}_l} \theta_l(w_l - t d_{l,n}, \widetilde{w}_l - t\tilde{d}_{l,n}) \right| \right) \\
&\leq \sum_{l=1}^{L} \sup_{\mathbf{t}\in[-1,1]^2} \left(\left| \frac{\partial^2}{\partial w_l^2} \theta_l(w_l - t_1, \widetilde{w}_l - t_2) \right| + \left| \frac{\partial^2}{\partial \widetilde{w}_l^2} \theta_l(w_l - t_1, \widetilde{w}_l - t_2) \right| \right. \\
&\qquad \left. +2 \left| \frac{\partial^2}{\partial w_l \partial \widetilde{w}_l} \theta_l(w_l - t_1, \widetilde{w}_l - t_2) \right| \right) \qquad (8.18)
\end{aligned}
$$

where we have used the fact that $|r_{n,l}|, |d_{n,l}|, |\tilde{d}_{n,l}| \leq 1$, for $n = 1 \ldots N$ and $l = 1 \ldots L$, since we are considering unit-norm filters. In a J-band undecimated filterbank, the upper-bound derived in Equ. (8.18) can be further tighten in a subband-adaptive manner, since $|r_{n,l}| \leq 1/M_j$, for all $n = 1 \ldots N$ and $(j-1)N + 1 \leq l \leq jN$ (the total number of transformed coefficients being usually given by $L = J \times N$). In the second row of Figure 8.2, we have plotted the values of the upper-bound (8.18) as a function of the average energy of the noise-free signal. In this particular case, the argument of the supremum has been computed for a discrete set of values $\mathbf{t} \in [-1,1] \times [-1,1]$ and we have then retained the maximum value. Although feasible, this first approach turns out to be sensitive to the computation of the supremum.

Based on the value of this supremum, it becomes then difficult to find an appropriate image-independent "threshold of reliability", above which, a non-linearly processed subband should be disregarded.

In practice, we suggest to consider an alternative empirical approach, based on the following observation: we notice in the first row of Figure 8.2 that the SNR of the approximation is a nearly linear (in a logarithmic scale) function of the average signal energy. We further observe that, for a given average signal energy, the quality of the approximation increases as the iteration depth j increases. Consequently, we suggest to only keep those non-linearly processed subbands $\mathbf{F}_{j,2}(\mathbf{y})$ for which $T_j = M_j E_{\text{mean}}$ is above a given threshold T; we experimentally found that any value of $T \in [5, 15]$ was a reasonable choice (see the third row of Figure 8.2). Thanks to this restriction, the PURE-based parameters optimization gives similar PSNR results to the MSE-based optimization, even when the latter considers all the non-linearly processed subbands $\mathbf{F}_{j,2}(\mathbf{y})$, as shown in Figure 8.3.

8.5 Experiments

In this section, we propose to compare our PURE-LET thresholding (8.15) applied in the Haar UWT with three multiscale-based methods in simulated experiments. We have retained the same algorithms and followed the same protocol as in Section 7.6.

8.5.1 Poisson Noise Removal

Before comparing our approach with the other state-of-the-art methods, we point out in Figure 8.4 that a PURE-LET thresholding applied in an undecimated Haar wavelet representation outperforms (over +1 dB) the *non-redundant* interscale thresholding (7.10) presented in Chapter 7. When several cycle-spins of the latter are considered, the gain ($0.3 - 0.5$ dB) becomes less important.

In Table 8.1, we compare our method with the retained state-of-the-art multiscale denoising algorithms. As observed, we obtain, on average, the highest PSNR results. The *TI-Haar-Fisz+TI-SUREshrink* algorithm is consistently outperformed by both our method and the *Platelet*. We can notice that the *GAT+BLS-GSM* solution becomes competitive for averaged intensities I_{mean} above ~ 10, whereas the other approaches are not restricted to high intensities images. We obtain similar results to the *Platelet* method, which was shown to be competitive with the recent approach described in [159]. However, the major drawback of the *Platelet* algorithm is its computation time: it may require around 1300s to denoise a single 256×256 grayscale image. To compare with, the execution of the *GAT+BLS-GSM* and of the *TI-Haar-Fisz+TI-SUREshrink* algorithm takes respectively ~ 7.4s and ~ 2.4s, whereas running the proposed solution only requires ~ 1.4s, under the same conditions.

In Figure 8.5, we show a visual result of the various algorithms applied on the *Moon* image. We can notice that the PURE-LET denoised image exhibits very few

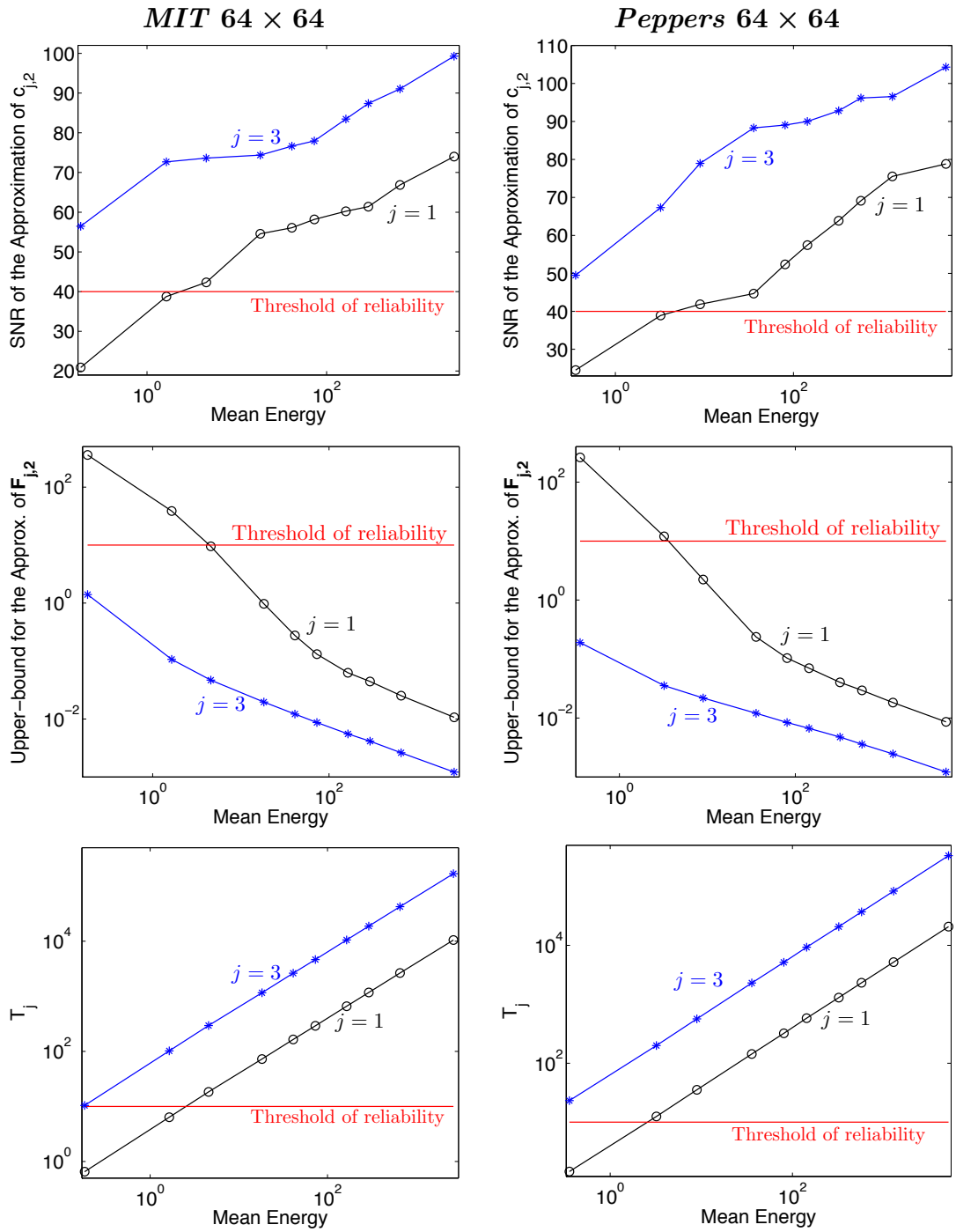


Figure 8.2: Validity of the first order Taylor series approximation of some nonlinear functions $\mathbf{F}^{-}_{j,2}$. The first row shows the SNR of the approximation of the component $c_{j,2}$ of the PURE vector $\mathbf{c}$. The second row displays the upper-bound on the Taylor series remainder of $\mathbf{F}^{-}_{j,2}$. In the third row are plotted the values of $T_j = M_j E_{\text{mean}}$.

Table 8.1: Comparison of multiscale-based Poisson noise removal algorithms

Images	***MIT*** 256×256						
$\mathbf{I_{mean}}$	38.00	19.00	9.50	6.33	3.17	1.58	0.32
Input PSNR	25.81	22.77	19.78	18.01	14.98	11.97	4.99
TI-Haar-Fisz+ TI-SUREshrink	29.71	27.72	25.25	23.84	22.66	20.89	16.82
GAT+BLS-GSM (FSP)	30.70	28.71	26.66	25.36	22.84	20.47	14.40
Platelet 25 CS	31.34	29.25	27.44	26.31	24.50	**22.96**	18.73
PURE-LET (8.15) **(UWT Haar)**	**31.52**	**29.43**	**27.50**	**26.44**	**24.57**	22.80	**18.82**
Images	***Cameraman*** 256×256						
$\mathbf{I_{mean}}$	56.52	28.26	14.13	9.42	4.71	2.36	0.47
Input PSNR	24.08	21.08	18.05	16.29	13.27	10.25	3.28
TI-Haar-Fisz+ TI-SUREshrink	29.44	27.35	25.19	24.54	23.08	21.53	18.98
GAT+BLS-GSM (FSP)	30.76	29.07	27.31	26.18	23.84	21.12	15.56
Platelet 25 CS	30.54	29.00	27.54	26.69	**25.33**	**23.56**	**20.53**
PURE-LET (8.15) **(UWT Haar)**	**31.03**	**29.29**	**27.67**	**26.72**	25.10	23.50	20.48
Images	***Fluorescent Cells*** 512×512						
$\mathbf{I_{mean}}$	21.56	10.78	5.39	3.59	1.80	0.90	0.18
Input PSNR	28.25	25.23	22.23	20.47	17.46	14.45	7.45
TI-Haar-Fisz+ TI-SUREshrink	34.63	32.99	31.47	30.37	28.69	27.42	24.42
GAT+BLS-GSM (FSP)	**35.17**	**33.35**	31.31	30.00	27.47	24.87	20.62
Platelet 25 CS	34.08	32.52	31.10	30.34	28.95	27.83	25.05
PURE-LET (8.15) **(UWT Haar)**	34.98	33.29	**31.72**	**30.84**	**29.41**	**28.10**	**25.16**
Images	***Moon*** 512×512						
$\mathbf{I_{mean}}$	34.08	17.04	8.52	5.68	2.84	1.42	0.28
Input PSNR	26.26	23.25	20.23	18.48	15.47	12.46	5.47
TI-Haar-Fisz+ TI-SUREshrink	29.63	27.60	25.41	24.71	23.87	23.20	22.13
GAT+BLS-GSM (FSP)	29.78	27.94	26.31	25.35	23.79	22.43	19.31
Platelet 25 CS	28.82	27.43	26.21	25.66	24.74	24.14	23.28
PURE-LET (8.15) **(UWT Haar)**	**29.98**	**28.35**	**26.95**	**26.24**	**25.21**	**24.47**	**23.52**

Note: Output PSNRs have been averaged over ten noise realizations, except for the *Platelet* algorithm.

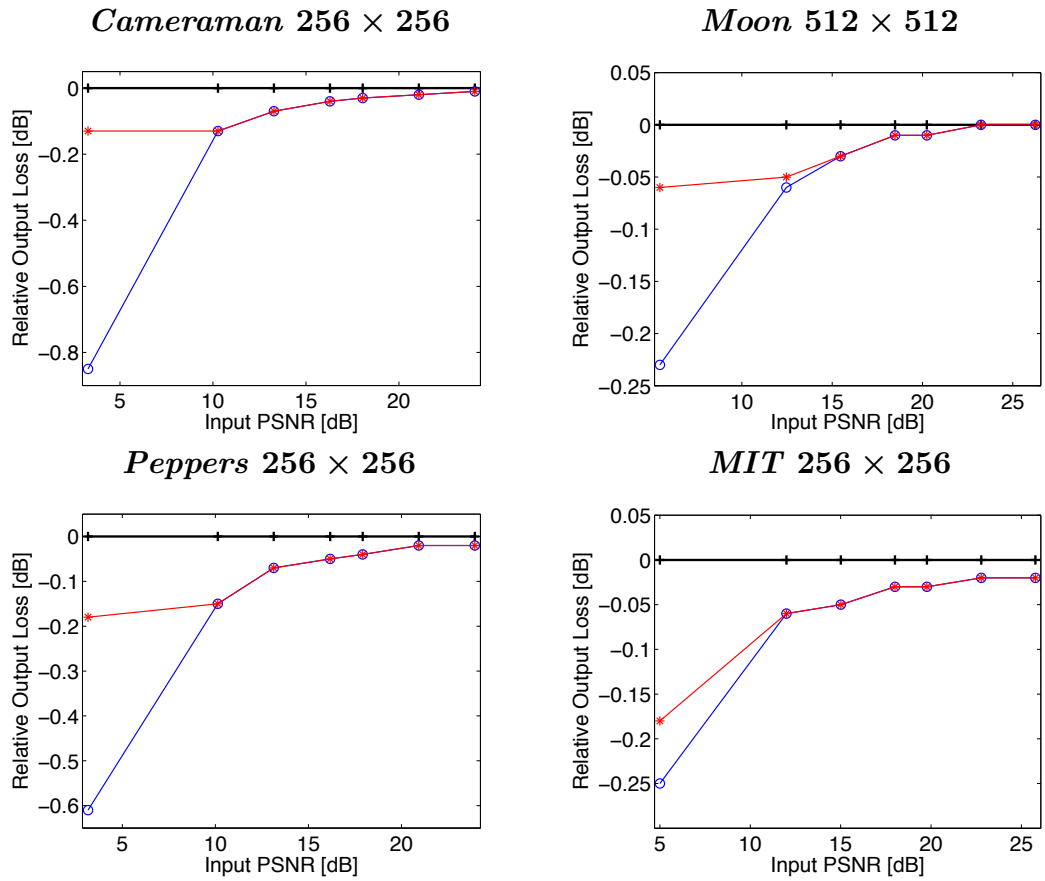


Figure 8.3: The blue curve ("○" markers) represents the output PSNRs obtained by a PURE-based parameters optimization without restriction, i.e. considering all the nonlinear terms. The red curve ("∗" markers) displays the output PSNRs obtained by a restricted PURE-based parameters optimization, i.e. all the non-linearly processed subband $\mathbf{F}_{j,2}(\mathbf{y})$ with $T_j < 10$ have been disregarded. The output PSNRs obtained by the MSE-based parameters optimization (without restriction) is the baseline ("+" markers).

artifacts, without over-smoothing, contrary to the other methods.

8.5.2 Poisson and Gaussian Noise Removal

Figure 8.6 shows the PSNR results obtained by *GAT+BLS-GSM* and by the proposed algorithm, when estimating Poisson intensities further degraded by an AWGN. Here again, our PURE-LET approach gives the best results. In particular, at low intensities, the GAT fails to stabilize the noise variance and thus, huge gains (> 1 dB) can be obtained with a direct handling of Poisson statistics. A visual comparison

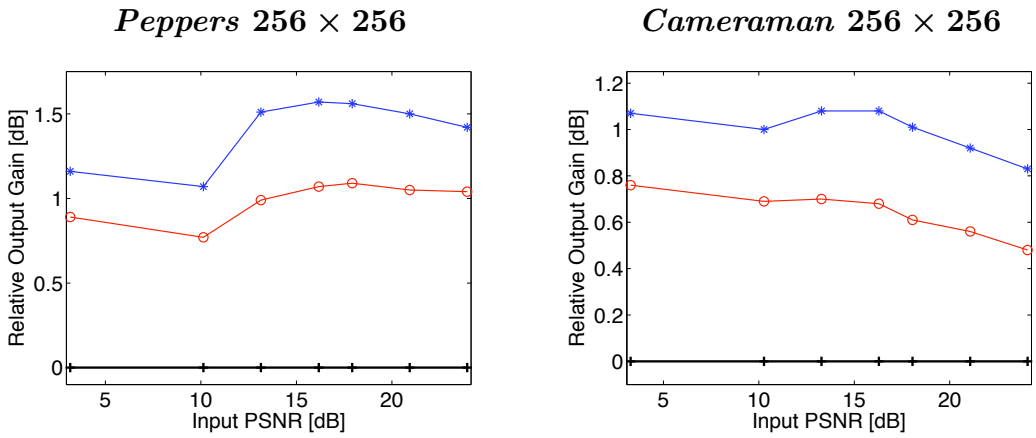


Figure 8.4: PSNR improvements brought by the redundant PURE-LET thresholding 8.15 ("∗" markers). The baseline ("+" markers) is provided by the non-redundant interscale thresholding (7.10). The red curve ("○" markers) represents the output PSNRs obtained by averaging 10 cycle-spins of the non-redundant interscale thresholding (7.10).

of both algorithms is given in Figure 8.7 for the standard *Barbara* image. As in the AWGN case, it is more advisable to apply the proposed PURE-LET algorithm with a 12×12 BDCT for this kind of textured image. Note that a dictionary-based representation is also conceivable, in order to get the best performances out of several complementary transforms.

8.6 Summary

In this chapter, the restricted use of the unnormalized Haar wavelet transform has been lifted: the PURE-LET strategy has been extended to arbitrary (redundant) transform-domain processing. In order to obtain a computationally fast and efficient algorithm, we have proposed a first order Taylor series approximation of PURE. For each non-linearly processed subband, the reliability of this approximation can be controlled. We have shown that, provided that a data-dependent quantity remains above a given threshold, the PURE-based parameters optimization achieves near *oracle* performances.

As an illustrative algorithm, we have proposed a simple subband-dependent thresholding applied to the coefficients of a J-band undecimated filterbank. In each bandpass subband, the amount of shrinkage is related to the signal-dependent noise variance, estimated from the corresponding lowpass coefficients. The resulting denoising algorithm has a reasonable computational complexity and is faster than other state-of-the-art approaches which make use of redundant transforms. It gives

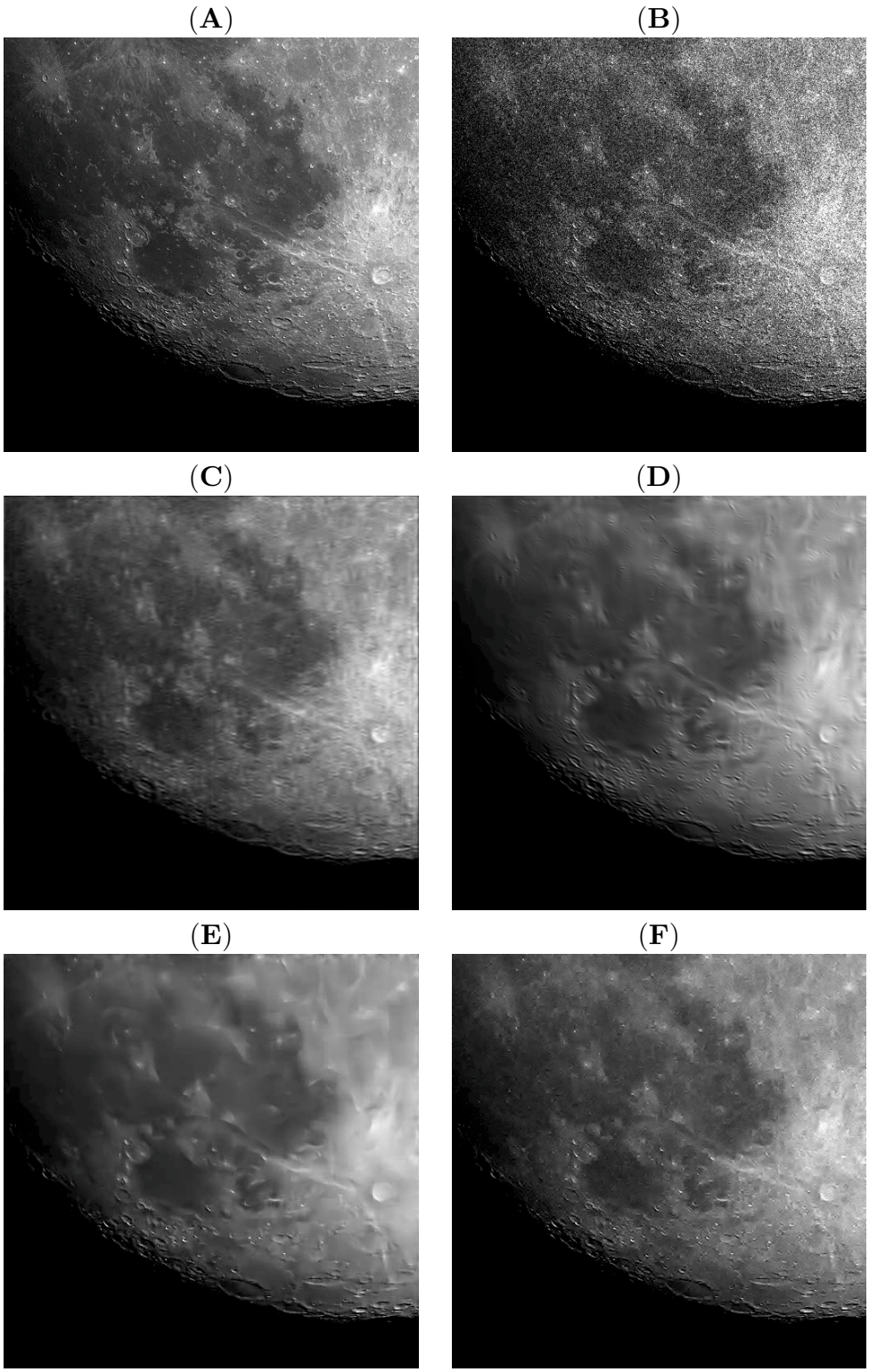


Figure 8.5: (**A**) The original 512×512 *Moon* image. (**B**) Noisy observation of it: PSNR = 18.48 dB. (**C**) Denoised with *TI-Haar-Fisz+TI-SUREshrink*: PSNR = 24.71 dB in 8.5s. (**D**) Denoised with *GAT+BLS-GSM*: PSNR = 25.31 dB in 39.3s. (**E**) Denoised with *Platelet*: PSNR = 25.66 dB in 3500s. (**F**) Denoised with the proposed algorithm: PSNR = 26.27 dB in 8.4s.

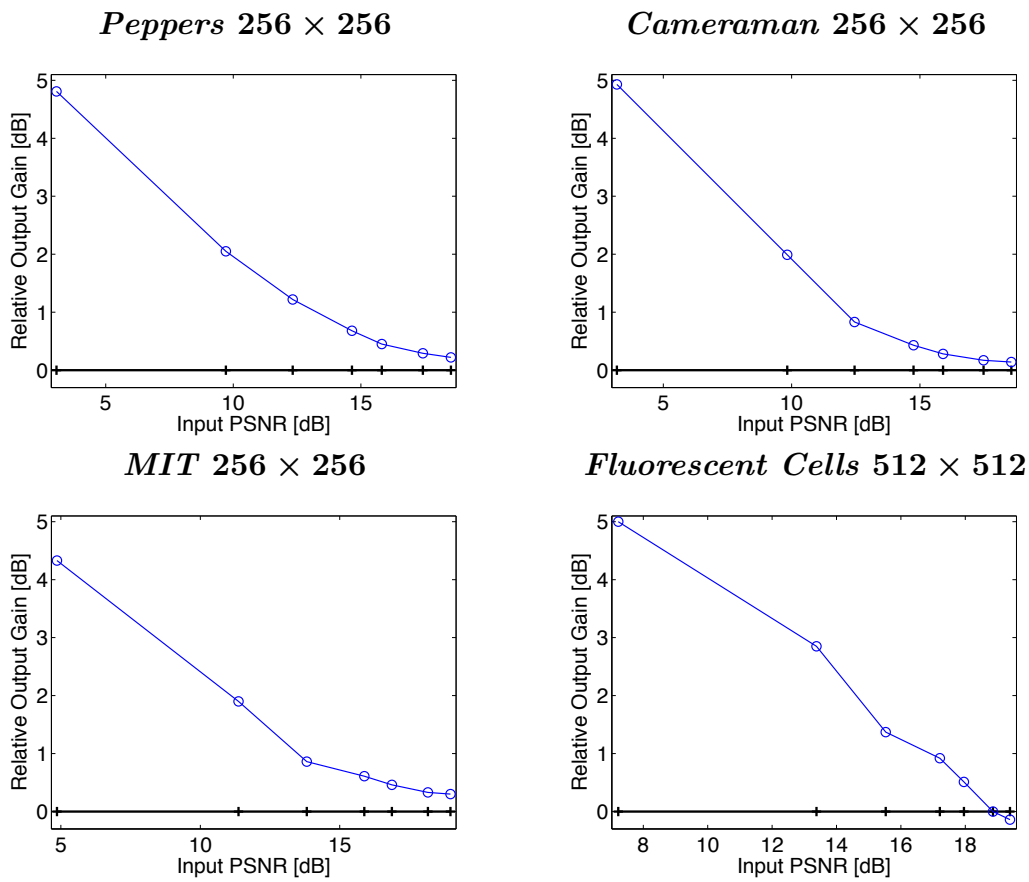


Figure 8.6: Comparison between the proposed Haar UWT *PURE-LET* thresholding (8.15) ("○" markers) and *GAT+BLS-GSM* applied in a FSP ("+" markers, benchmark).

better results, both quantitatively (lower MSE) and qualitatively, than the standard VST-based algorithms. The proposed solution also favorably compares with one of the most recent multiscale methods specifically devised for Poisson data.

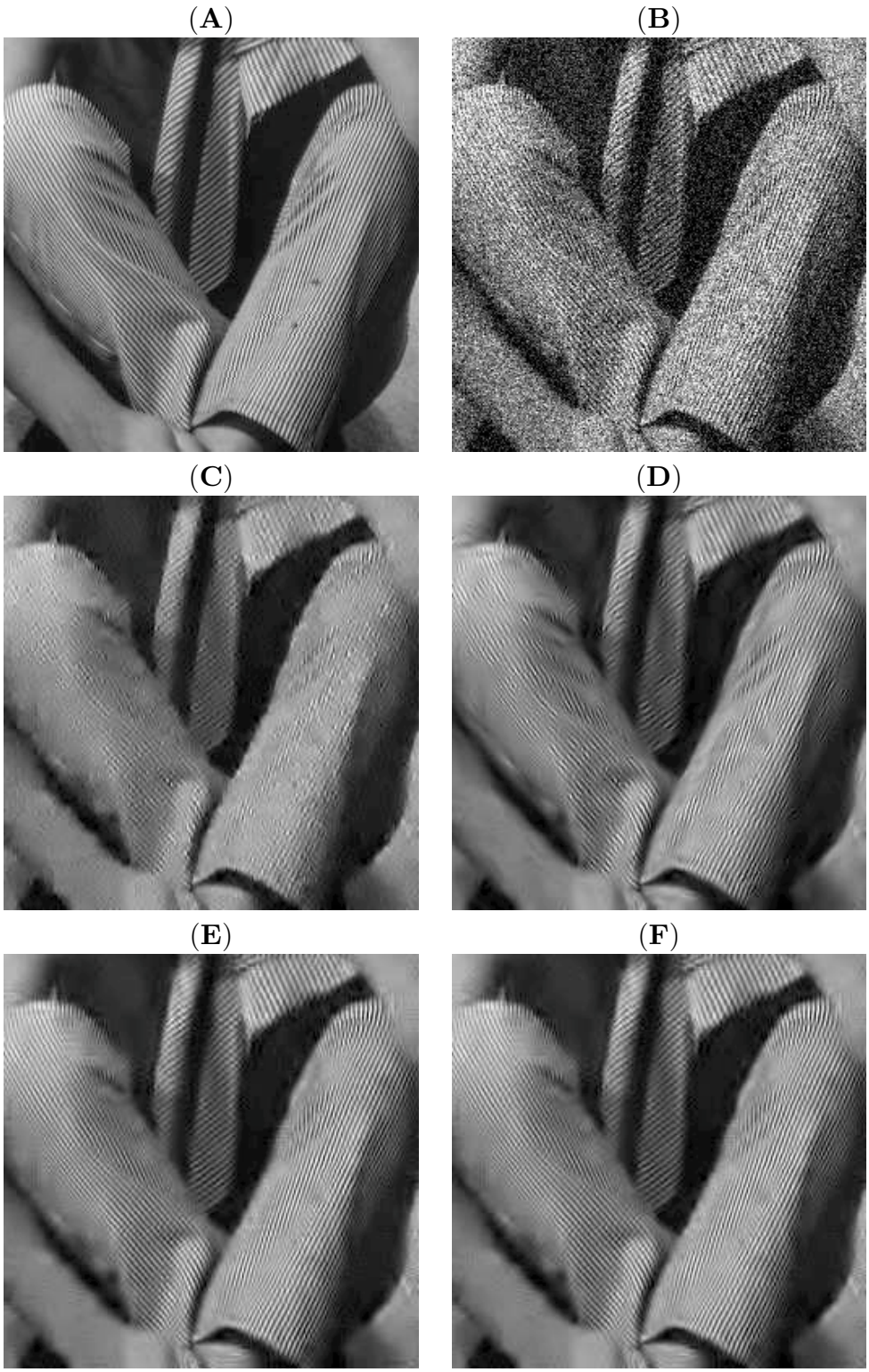


Figure 8.7: (**A**) Part of the original *Barbara* image. (**B**) Noisy observation of it: PSNR = 15.87 dB. (**C**) Denoised with UWT Haar PURE-LET: PSNR = 24.62 dB.(**D**) Denoised with GAT+*BLS-GSM*: PSNR = 26.04 dB. (**E**) Denoised with 12×12 BDCT PURE-LET: PSNR = 26.87 dB. (**F**) Denoised by PURE-LET optimized in a UWT/BDCT dictionary: PSNR = 26.93 dB.

Chapter 9

Denoising of Fluorescence Microscopy Data

9.1 Motivations

The incessant development of improved microscopy imaging techniques, as well as the advent of highly selective fluorescent dyes has made possible the precise identification of tagged molecules in almost any biological specimen. Of particular interest are the visualization and the study of *living* cells, which induce tight constraints on the imaging process. To avoid the alteration of the sample and to achieve a high temporal resolution, low fluorophore concentrations, low-power illumination and short exposure time need to be used in practice. Such restrictions have a tremendous impact on the image quality. This is why the denoising of fluorescence microscopy images is currently an active area of research [169,171–174], the main challenge being to provide efficient, fast and automatic denoising algorithms, capable of processing huge multidimensional biomedical datasets.

Under the imaging constraints previously evoked, the intensity of the measured fluorescent radiations can be very low leading to highly signal-dependent random fluctuations. A common assumption is then to consider that the noisy measurements are the realizations of a Poisson process, possibly further degraded by an AWGN. In this chapter, we first justify the adequacy of this common statistical model in real fluorescence microscopy data. Secondly, we assess the performance of the proposed PURE-LET denoising strategy from a quantitative, qualitative and practical (computation time) point of view.

9.2 Principle of Fluorescence Microscopy

In this section, we discuss the two key ingredients involved in fluorescence microscopy: the fluorophore and the light microscope, either wide-field or confocal. For a more detailed introduction to fluorescence microscopy, we refer the reader to [142] and to the websites of the microscope manufacturers[1].

A fluorophore is a molecule that can emit lights at a given wavelength, when illuminated by a source of light of a specific, shorter wavelength. The green fluorescent protein (GFP) discovered by Shimomura *et al.* in the early sixties, is probably the most popular fluorescent dye. This intrinsically fluorescent protein is naturally produced by the *Aequorea victoria* jellyfish. It can be linked to a target protein and then introduced in given cells or embryos. The latter will finally synthesized the labeled protein.

To observe the molecule of interest, the sample is placed under the objective of a light microscope. The illuminating source (typically located in the blue range of the spectrum) excites the specimen, and the light (usually in the green part of the spectrum) reemitted by the labeled proteins is collected by the microscope sensors. In standard wide-field microscopy, the illumination is provided by a powerful arc-discharged lamp and the detection is performed by a charge-coupled device (CCD). In confocal microscopy, a laser is used as the light source and the detection is usually operated by a photomultiplier tube (PMT), only occasionally by CCD sensors. The full acquisition process of a confocal microscope is shown in Figure 9.1. The main difference between this type of microscope and the conventional wide-field microscope is the presence of a pinhole which can reject the out-of-focus light, thus increasing the axial resolution. Since one point is imaged at a time, the blurring effect is highly reduced. However, since less light is collected, the signal-to-noise ratio is usually lower in confocal microscopy. To acquire the whole sample, a scanning unit is required to move the focused laser beam across the specimen. The xy-scanning is usually performed line-by-line, while the axial z-position is adjusted by a motorized high-precision table.

In light microscopy, the resolution is mainly limited by diffraction. Theoretically, it is proportional to the wavelength of the light emitted by the fluorophores and inversely proportional to the numerical aperture of the objective lens. In confocal microscopy, the axial resolution is approximately three times that of the lateral resolution. In practice, the lateral resolution of a standard light microscope can rarely go below $200nm$.

9.3 Noise in Fluorescence Microscopy

We now briefly review the main sources of noise in fluorescence microscopy. For a more detailed investigation, the reader can refer to [142, 175].

[1] For instance, the following website: `http://www.microscopyu.com/`

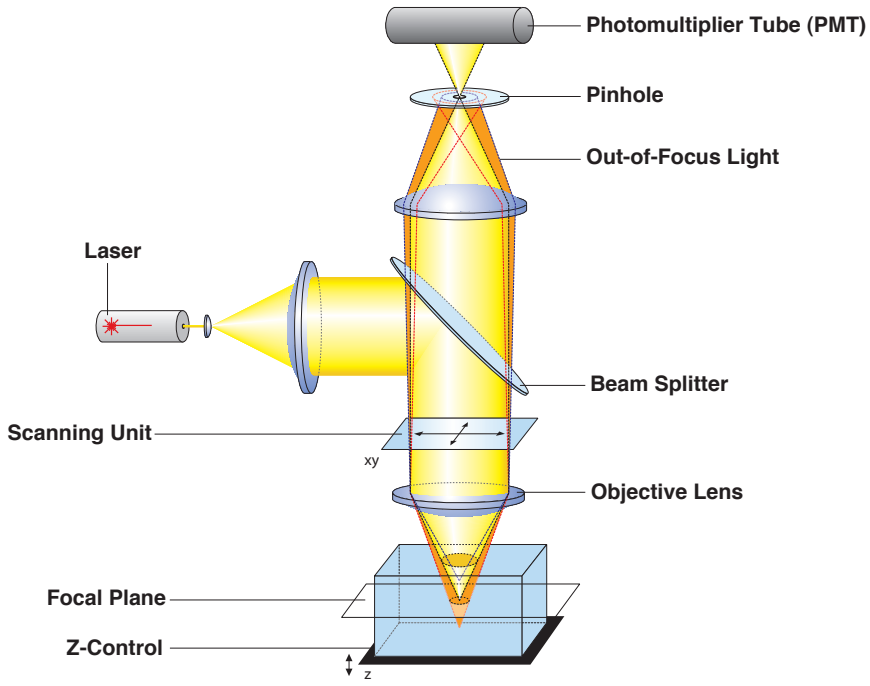


Figure 9.1: Schematic representation of a confocal microscope (extracted from a *Zeiss* datasheet and re-editated).

9.3.1 Photon-Counting Noise

Photon-counting noise is probably the major source of noise in fluorescence microscopy. Indeed, light emission, whether during illumination, observation or detection, is a random process where photons are emitted at random time intervals. This inherent statistical variation in the emission rate of photons is well-described by a Poisson process.

In confocal microscopy, the measurement process is the same for every scan position (pixel); it is illustrated in Figure 9.2 and can be summarized as follows. The random number of photons arriving at the PMT during the integration time follows a Poisson distribution of mean λ determined by the source intensity. Each photon may traverse the protection window, penetrate the photocathode and be converted to an electron with a certain probability η; this probability is known as the quantum efficiency of the PMT and is on the order of 30% for the best models. The conversion process can be seen as a binomial selection [176, 177] and thus, the number of electrons at the output of the photocathode (photoelectrons) follows a Poisson distribution of mean $x = \eta\lambda$. The number of photoelectrons represents a shot-noise-corrupted measurement y of the intensity signal x in our framework.

The electrons are then multiplied (via several amplification stages) and converted

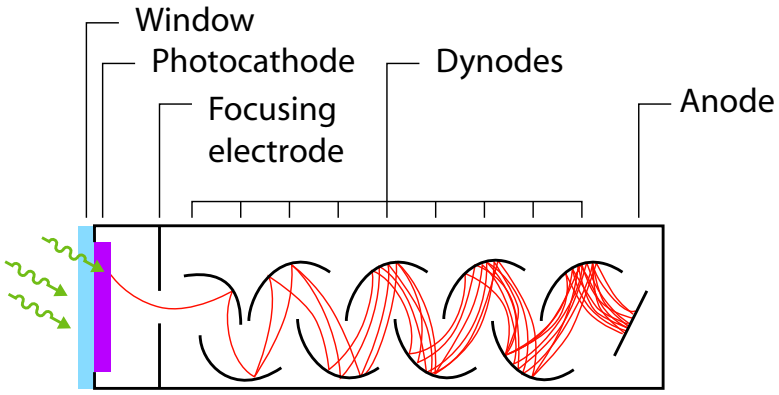


Figure 9.2: Schematic representation of a photomultiplier tube (PMT).

to an electric current that is integrated and quantized. The recorded signal is essentially proportional to the number of photoelectrons; although the amplification factor may fluctuate in practice, recent work [178] suggests that the newest PMTs with high first-stage gain have the ability to discriminate between multi-photoelectron events, at least for low count numbers.

9.3.2 Measurement Noise

In addition to photon counting noise, signal-independent noise contributions also exist due to the imperfection of the electronic devices and the analog to digital conversion. In particular, even when no signal is transmitted, the output of the detectors might fluctuate around a non-zero value, known as the *dark current*. These fluctuations are mainly due to the intrinsic thermal instabilities of the detectors. They can be reduced by cooling the detectors. When all put together, these signal-independent noise contributions can be considered as normally distributed.

9.3.3 Sample-Dependent Noise

Finally, one can also encounter two main sources of signal degradation that come from the sample itself.

The first is due to potential intrinsic fluorescence properties of the specimen. Some non-labeled molecules can also emit fluorescence radiation which may interfere with that of the tagged molecule, a phenomenon coined *autofluorescence*. Autofluorescence alters the extraction of the signal of interest if the emission wavelengths of the non-labeled molecules overlap with that of the tagged molecules.

The second comes from the loss of fluorescence abilities of the fluorophore due to chemical reactions induced by the source of illumination. This phenomenon, called *photobleaching*, leads to a significant drop of signal intensities. It can be reduced by limiting the exposure time of the fluorophores, lowering the energy of the illuminating source and/or increasing the fluorophores concentration. Nevertheless, these techniques also reduce the number of detected photons, and consequently, decrease the SNR.

9.4 Noise Parameters Estimation

To account for the various sources of noise described in Section 9.3, we propose to consider the following (simplified) noisy measurements model: a scaled Poisson law degraded by an AWGN with a possibly non-zero mean, i.e.

$$y \sim \alpha \mathcal{P}(x) + \mathcal{N}(\delta, \sigma^2) \tag{9.1}$$

where α and δ respectively model the gain and the offset of the detectors.

The knowledge of the parameters $(\alpha, \delta, \sigma^2)$ is essential for the calibration of our noisy measurements. δ and σ^2 may be measured by running separate calibration experiments [179] or by computing the sample-mean and sample-variance in signal-free (background) regions of the noisy image. The parameter α is usually not equivalent to the gain factor that is typically provided in microscopes and therefore, it must be estimated from the given data. For this, we devise a simple mechanism to estimate α and $(\sigma^2 - \alpha\delta)$ based on the following identities:

$$\mu_y \overset{\text{def}}{=} \mathscr{E}\{y\} = \alpha x + \delta \tag{9.2}$$

$$\sigma_y^2 \overset{\text{def}}{=} \text{Var}\{y\} = \alpha^2 x + \sigma^2 \tag{9.3}$$

This leads to the relationship:

$$\sigma_y^2 = \alpha \mu_y + \underbrace{(\sigma^2 - \delta\alpha)}_{\beta} \tag{9.4}$$

μ_y and σ_y^2 are robustly estimated by computing the sample-mean and sample-variance inside every non-overlapping $M \times M$ (usually 8×8) blocks of the image. Once that is done, we simply perform a robust linear regression (e.g. an iterative weighted least-square fit) on the set of points (μ_y, σ_y^2): the slope yields an estimate of the gain α and the intercept at $\mu_y = 0$ yields an estimate of β. The robustness of the estimation can be increased by repeating this simple procedure to several cycle-spins of the input image, and then averaging the parameters estimated for each cycle-spin. Note that a similar procedure was already proposed by Lee *et. al.* in 1989 for the estimation of the variances of a multiplicative and an additive noise [97], and more recently by Boulanger in [180].

An example on simulated noisy data (*Fluorescent Cells*) ($\alpha = 5, \delta = 120, \sigma^2 = 16$) is provided in Figure 9.3. The parameters have been accurately estimated as ($\hat{\alpha} = 4.99, \hat{\delta} = 121.58, \hat{\sigma}^2 = 16.81$).

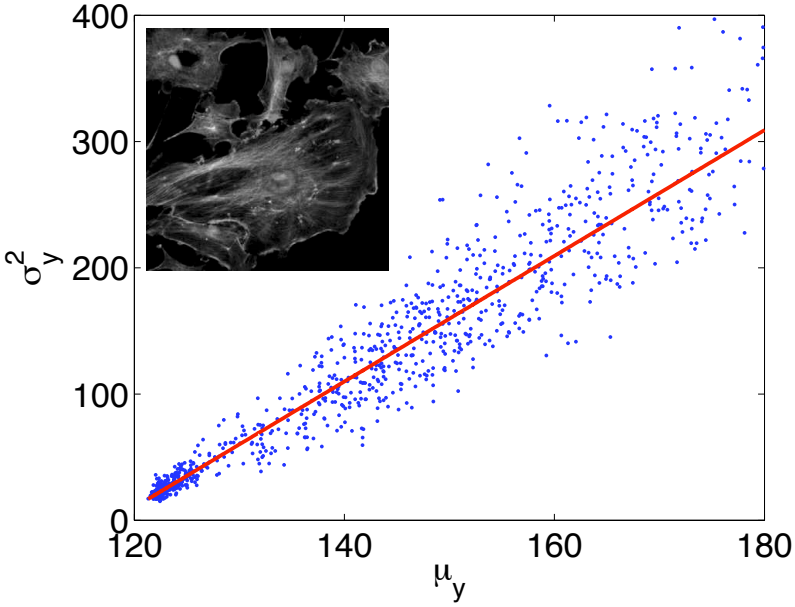


Figure 9.3: Automatic noise parameters estimation.

9.5 Experimental Results

9.5.1 2D Fixed Sample

In this section, we want to validate the statistical observation model and to assess the denoising quality of the proposed PURE-LET approach on fluorescence-microscopy images. The two considered datasets contain 100 images of a fixed sample acquired under low illumination intensity, i.e. mainly shot-noise limited. For these datasets, we thus assume that each noisy pixel y is distributed according to

$$y \sim \alpha \mathcal{P}(x) + \delta$$

First Dataset

The first dataset contains 100 3-channel 512×512 images acquired on a Leica TCS SP5 confocal microscope at the Imaging Center of the IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France). This microscope is equipped with a water-cooled scan-head and low-noise photomultiplier tubes

(PMTs). The fixed sample is a *C. elegans* embryo labeled with three fluorescent dyes (Figure 9.6(A)). Very noisy experimental conditions were reproduced by reducing the laser power and using short exposure times. The three color channels were processed separately.

The parameters α and δ have been estimated by the procedure described in Section 9.4. In Figure 9.4, we show the computed mean/variance couples (μ_y, σ_y^2) and the corresponding robust linear regression.

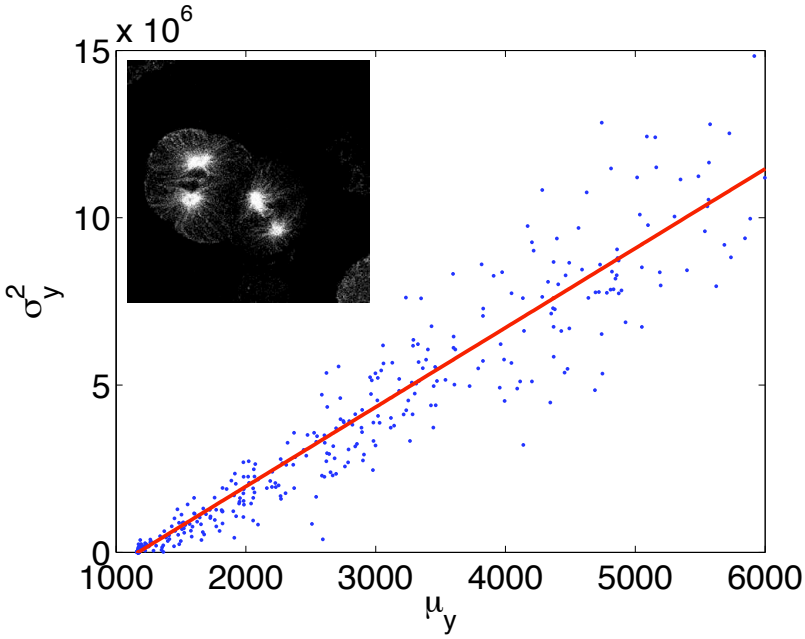


Figure 9.4: Automatic noise parameters estimation for the green channel of one particular image from the first fluorescence-microscopy dataset.

After subtracting the estimated detector offset δ and dividing the input data by the estimated gain α, the frequency distribution of the pixel values was found to be in good agreement with Poisson statistics. Figure 9.5 shows the histogram of the normalized pixel values of the green channel for those pixels whose mean was equal to a given value x ($\pm 5\%$).

In Figure 9.6, we show the denoising result obtained by the proposed solution and by the *TI-SUREShrink* algorithm with a *TI-Haar-Fisz* variance-stabilizing transform (Figure 9.6(C)). The result of our PURE-LET algorithm, either applied in a non-redundant representation (Figure 9.6(D)) or using the undecimated Haar wavelet (Figure 9.6(E)), is sharper and shows less artifacts. Besides, it is less noisy than the image shown in Figure 9.6(B), which corresponds to a 4 times longer exposure time.

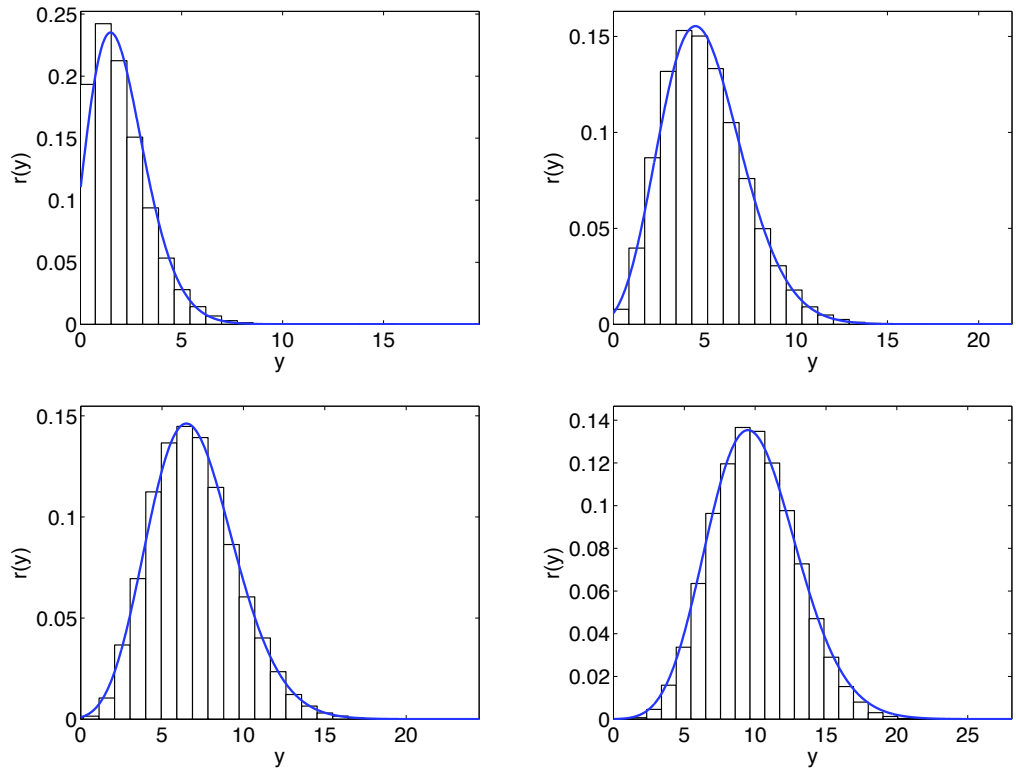


Figure 9.5: Comparison of the Poisson distribution with the histogram of the normalized data for underlying intensities $x = 2, 5, 7, 10$.

Second Dataset

The second dataset contains 100 512×512 images of fixed GFP-labeled tobacco cells acquired on a Leica SP5 confocal laser scanning microscope at the Department of Biology of the Chinese University of Hong Kong[2]. The noise parameters (α, δ) have been estimated as $\hat{\alpha} = 1087.04$ and $\hat{\delta} = 792$ (see Figure 9.7).

Since the tobacco cells were perfectly fixed, we can assume that the averaged of the 100 images constitutes a reliable estimate of the underlying true Poisson intensities. Considering this averaged image as the ground truth allows us to compute the signal-to-noise ratio (SNR) of any image of the dataset, and also compared various denoising algorithms. In Table 9.1, we have thus reported the SNR values obtained by the proposed PURE-LET methods, as well as those achieved by a standard median filter (with optimal support size) and by some state-of-the-art approaches. As observed our *non-redundant* PURE-LET (7.10) significantly outperforms the simple

[2]We are especially grateful to Prof. Liwen Jiang and Dr. Yansong Miao for the preparation of the biological sample and their help on acquiring the images.

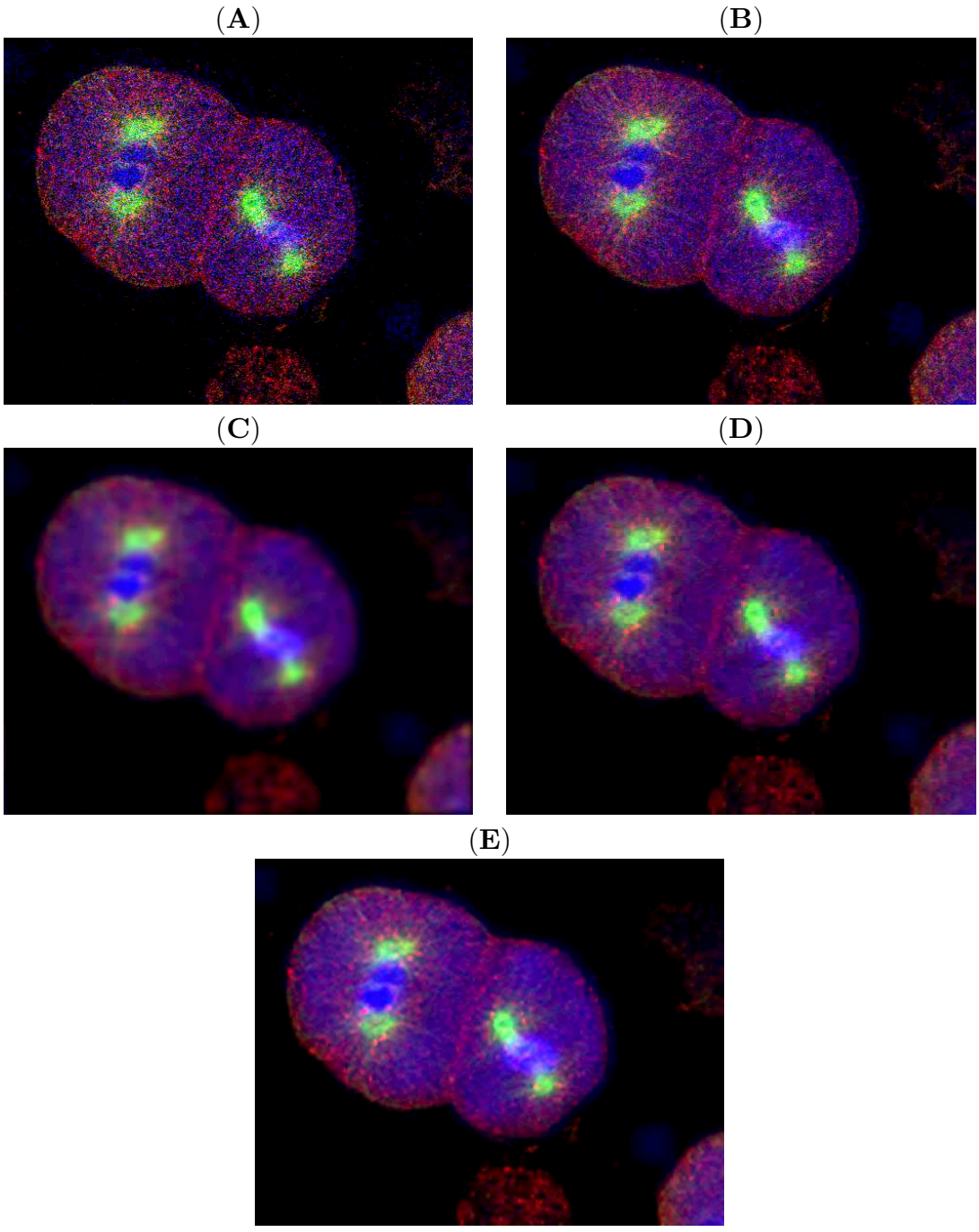


Figure 9.6: (**A**) Raw input image (scan frequency of 1000 Hz). (**B**) Average of 4 images at 1000 Hz). (**C**) Input image denoised with *TI-Haar-Fisz* + *TI-SUREshrink*. (**D**) Input image denoised with our non-redundant inter-scale *PURE-LET* (7.10). (**E**) Input image denoised with our undecimated Haar wavelet *PURE-LET* thresholding (8.15).

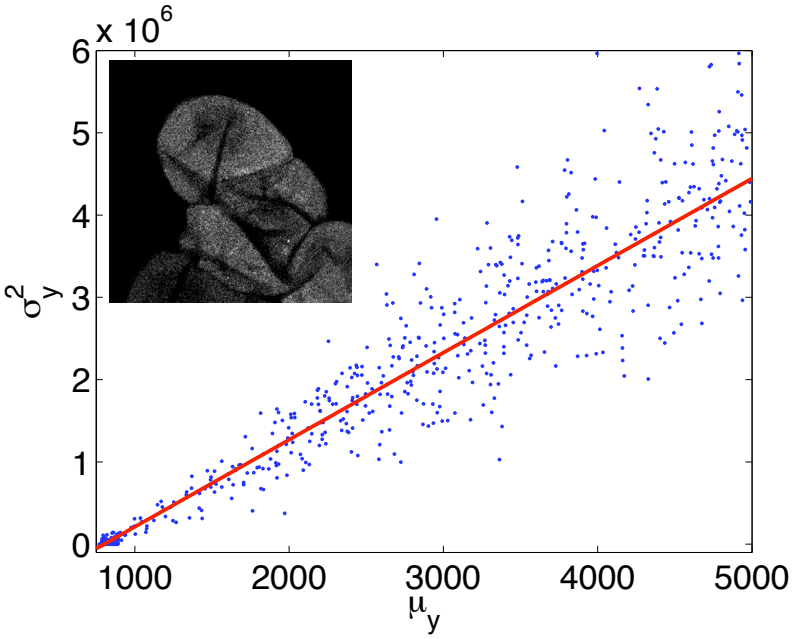


Figure 9.7: Automatic noise parameters estimation for one particular image from the second fluorescence-microscopy dataset.

median filter (+2 dB), a sophisticated AWGN denoiser combined with a standard variance-stabilizing transform (+1.2 dB), and the *non-redundant* variant of the state-of-the-art *Platelet* algorithm (+1 dB). The *redundant* variants of our PURE-LET strategy are competitive with the redundant *Platelet* method, while being one order of magnitude faster.

9.5.2 3D Sample

We acquired a $1024 \times 1024 \times 64$ stack of fluorescence images at the BioImaging and Optics platform (BIOP) at EPFL. We used a confocal microscope equipped with a 63X PL-APO objective. The X-Y pixel size was set to 0.09 $\mu m \times 0.09$ μm and the Z-step was $0.37\mu m$. In addition to fibroblast cells labeled with a DiO dye (which is predominantly retained by the cell membranes), the sample contained 100-nm fluorescent microbeads acting as point sources (see Figure 9.9(A)).

The noise parameters $(\alpha, \delta, \sigma^2)$ have been estimated as $\hat{\alpha} = 125.01$, $\hat{\delta} = 56.59$ and $\hat{\sigma}^2 = 11.72$ on the 30^{th} slice of the 3D stack (see Figure 9.8). Estimations performed on the other slices led to similar values for these parameters, confirming that the illumination conditions were identical for the whole stack.

In Figure 9.9, we display the denoising results of the standard 3D median filter (B), the recent *Platelets* approach (C) and the proposed algorithm (D). Observe

Table 9.1: SNR and computation time of various Poisson denoising algorithms applied on the *Tobacco Cells* fluorescence microscopy dataset.

***Tobacco Cells* at 4.59 dB**		
Method	**SNR [dB]**	**Computation Time [s]**
9 × 9 Median	14.82	2.15
GAT+BLS-GSM (FSP)	15.58	34.81
Platelet (1 CS)	15.89	41.83
PURE-LET (7.10) **(1 CS)**	16.77	**0.62**
PURE-LET (7.10) **(10 CS)**	17.33	6.03
Platelet (10 CS)	17.56	410
PURE-LET (8.15) **(UWT Haar)**	**17.65**	8.42

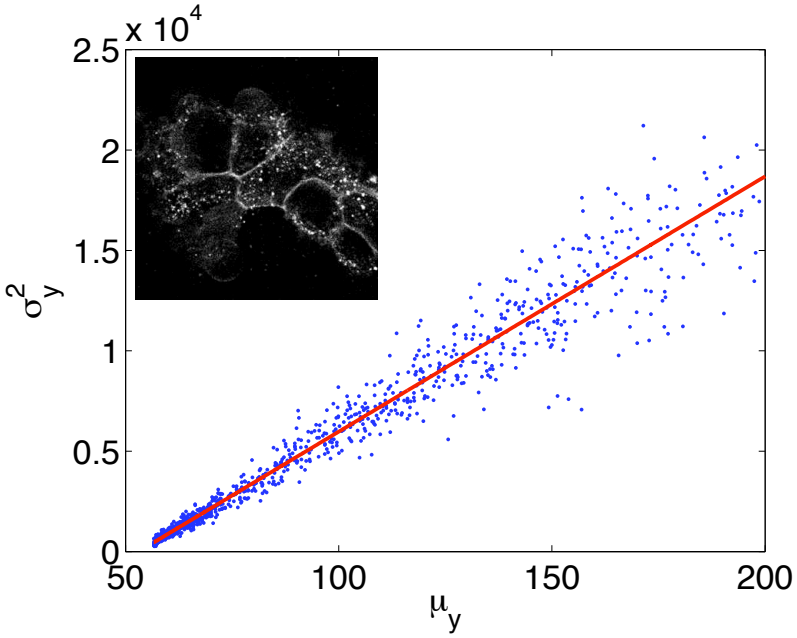


Figure 9.8: Automatic noise parameters estimation for one particular slice of the considered 3D fluorescence-microscopy stack.

that our solution compares favorably with the state-of-the-art *Platelets* technique, while being approximately three orders of magnitude faster.

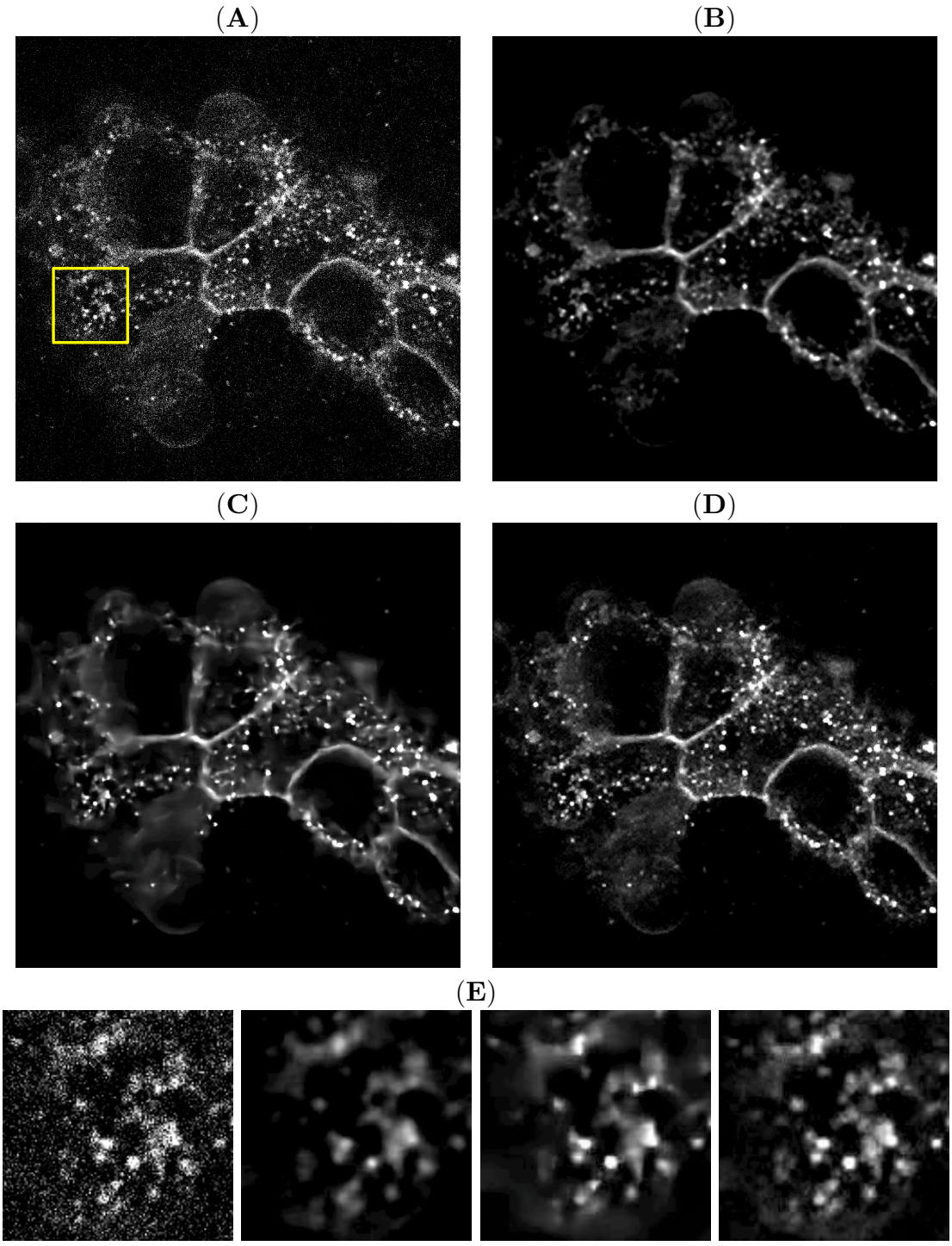


Figure 9.9: Part of a particular 1024×1024 slice of the 3D fluorescence stack. (**A**) Raw slice. (**B**) $7 \times 7 \times 3$ Median filter: 8.4s. (**C**) 25 cycle-spins of *Platelet*: 42min. (**D**) The proposed interscale multichannel *PURE-LET* (7.17) using $C = 3$ adjacent slices: 3.5s. (**E**) Zoom in images A, B, C and D.

9.5.3 2D Video Sequence

Finally, we show the result of the proposed PURE-LET denoising strategy applied on a real timelapse fluorescence microscopy image sequence[3] of *C. elegans* embryos[4]. The embryos were imaged with an Olympus IX 81 motorized inverted microscope equipped with a spinning disk Perkin Elmer Ultra View ERS. 1000 512×448 frames were imaged at 2 Hz for a total duration of 8min20sec. Frame number 47 is displayed in Figure 9.11(A).

In this special worm line, a GFP encoding cassette was inserted under the promoter of tubulin, a major cell cytoskeleton component active during cellular division. The green fluorophore thus tags the tubulin molecules in the embryos. More precisely, we see two *C. elegans* embryos at different stages of early development. The lower embryo has just gone through its very first division, whereas the upper embryo has already gone through multiple divisions.

With the estimation procedure described in Section 9.4 applied on the 47^{th} frame (see Figure 9.10), we obtained the following noise parameters: $\hat{\alpha} = 9.83$, $\hat{\delta} = 455.02$ and $\hat{\sigma}^2 = 164.86$. Here again, we noticed that all frames follow the same observation model.

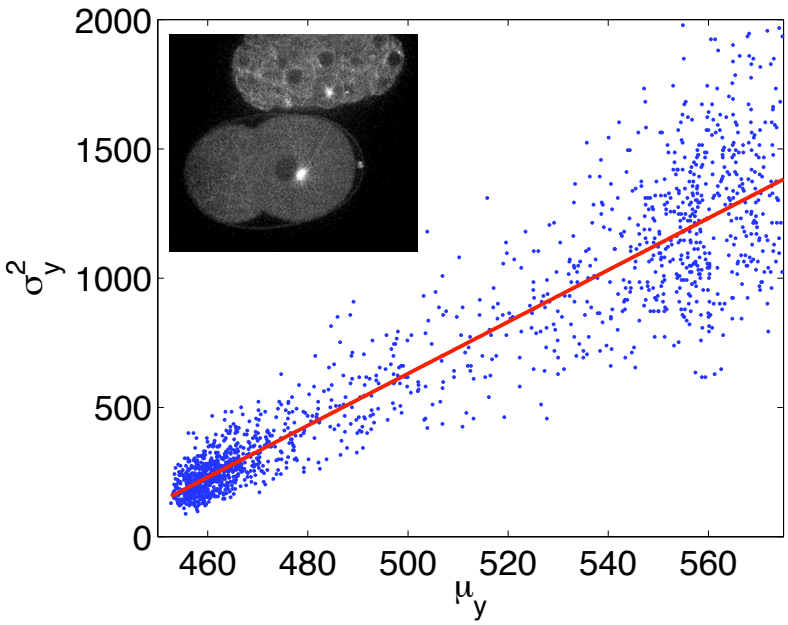


Figure 9.10: Automatic noise parameters estimation for one particular frame of the considered fluorescence-microscopy video sequence.

[3] Acquired with the invaluable help of the BIOP team at EPFL.
[4] Kindly provided by Prof. Pierre Gönczy, UPGON, ISREC, EPFL.

As can be seen in Figure 9.11, the denoising quality of the proposed interscale multichannel *PURE-LET* (7.17) (D) is much higher than a standard procedure such as the 3D median filtering (B). The proposed solution also exhibits fewer artifacts than the GAT followed by the sophisticated *BLS-GSM* applied in a full steerable pyramid (C), while being one order of magnitude faster. Many relevant details are extracted from the noisy image (A). Polar bodies that were ejected during meiosis get clearly visible as two different entities in the lower embryo, whereas as a single entity with the conventional 3D median filter. The linear shape of microtubules appears more clearly, in a star-like disposal from the centrosome to the periphery in the right cell (E) and globally dispersed in the left cell of the lower embryo, enabling both cells to be visually easily separable from one another. In the upper embryo, the various cells can now be distinguished from one another, with internal filamentous structure appearing in each cell.

9.6 Summary

In this chapter, we have briefly recalled the principle of fluorescence microscopy and discussed the various sources of noise inherent to this type of imaging technique.

We have then propose a simplified statistical measurement model that accounts for the signal-dependent and independent noise contributions. We have provided a simple and robust procedure to automatically estimate the underlying noise parameters. Such an observation model has been shown to properly reflects the statistics of real fluorescence measurements.

Finally, we have presented promising denoising results of 3D and timelapse fluorescence microscopy data. In particular, we have shown that the proposed PURE-LET approach was very competitive with the state-of-the-art, while being order(s) of magnitude faster. This latter point is really valuable in real applications, where large datasets have to be routinely processed.

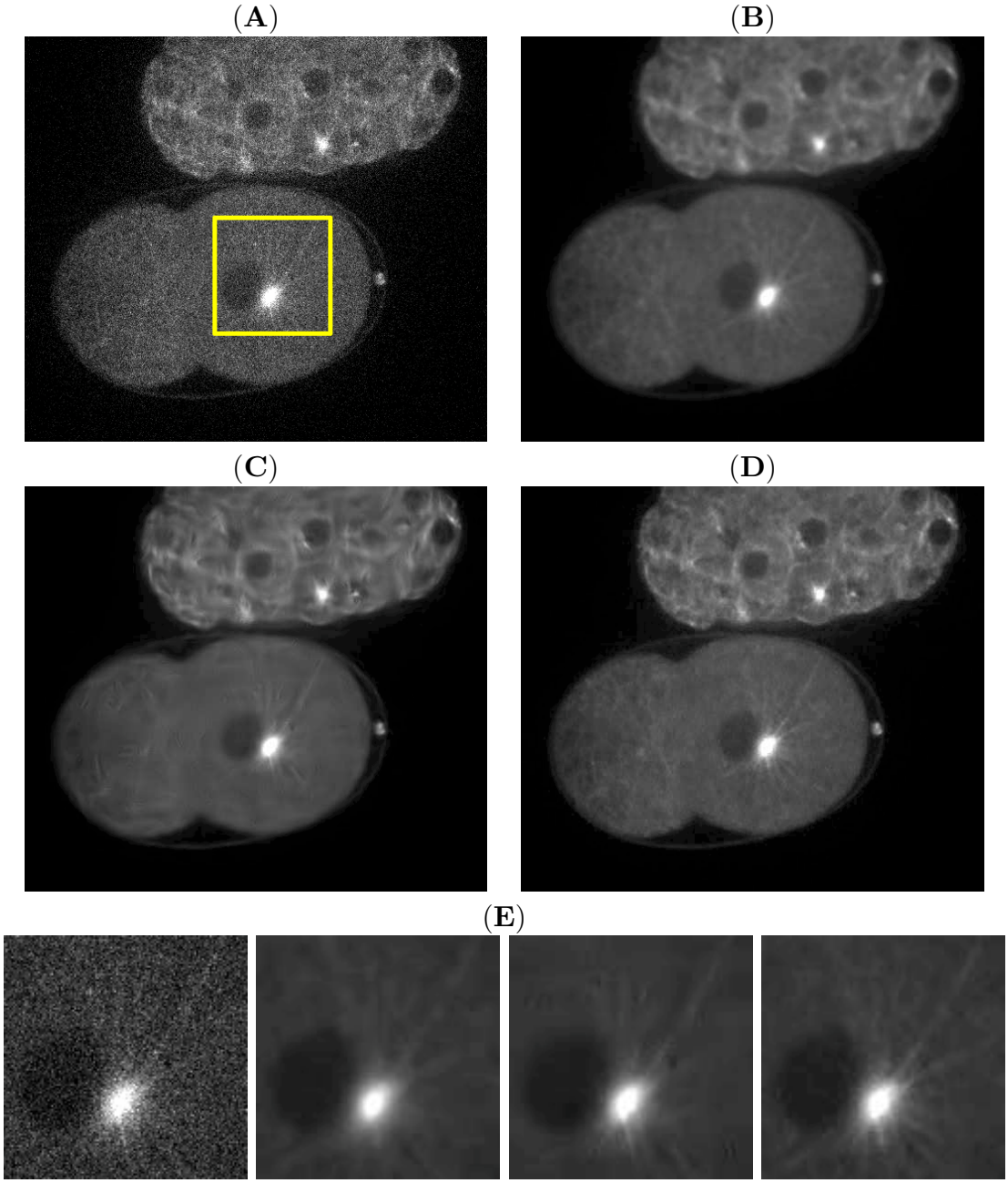


Figure 9.11: (**A**) Part of a particular 512×448 frame of the fluorescence microscopy image sequence. (**B**) $7 \times 7 \times 5$ Median filter: 9s. (**C**) *GAT+BLS-GSM* (FSP): 37s. (**D**) The proposed interscale multichannel *PURE-LET* (7.17) using $C = 5$ adjacent slices: 3s. (**E**) Zoom in images A, B, C and D.

Chapter 10

Epilogue

10.1 Summary

In this work, we proposed a novel *prior-free* approach for image denoising. In order to come up with fast and high-quality denoising algorithms, we devised a generic "URE-LET" framework. Indeed, the development of data-driven unbiased risk (expected MSE) estimates (UREs) allows to monitor the denoising quality, while a linear parameterization of the denoising process through a linear expansion of thresholds (LET) leads to a direct (i.e. non-iterative) parameters optimization.

In the first part of this thesis, we tackled the classical problem of additive white Gaussian noise (AWGN) reduction. In particular, we revitalized a relatively old statistical result, nowadays known as Stein's unbiased risk estimate (SURE). We worked out several SURE-based estimators applicable in arbitrary linear transformed domains. For non-orthogonal transforms, we pointed out the sub-optimality of a transform-domain optimization compared to a global image-domain SURE/MSE minimization. In the context of orthonormal wavelet denoising, we developed a rigorous procedure based on group delay compensation to take full advantage of the interscale dependencies that are naturally present in the wavelet representation of most images. We showed that a suitable integration of these interscale dependencies, combined with an efficient incorporation of the intrascale correlations of the wavelet coefficients, places the proposed SURE-LET estimators among the best state-of-the-art multiresolution-based denoising algorithms.

In the second part of this thesis, we considered a more realistic measurement model, where the observed noisy image is assumed to be the realization of a Poisson process degraded by an AWGN. Such a model accounts for the hybrid nature of real-world noise in images, which contains both a signal-dependent (due to the random nature of photon counting) and a signal-independent (due to the imperfections of the light detectors) component. In a Bayesian framework, it becomes much more difficult

to derive closed-formed estimators for this non-Gaussian observation model, whereas the proposed URE-LET paradigm is still applicable. Indeed, we derived a novel unbiased estimate of the MSE, coined *PURE*, that can be used to optimize a large class of Poisson intensity estimators (in particular of LET type). We showed that the (unnormalized) Haar wavelet transform is the only multiresolution transform where the global image-domain PURE can be split into independent subband-PUREs. We then proposed an interscale Haar wavelet domain PURE-LET estimator that turns out to be competitive with the standard VST-based denoisers applied in a *redundant* wavelet representation. For an arbitrary transform, the exact PURE becomes time-consumming to compute for non-linear processing. We thus proposed a reliable first order Taylor series approximation of PURE that can be used in practice. As an illustrative example, we devised a novel undecimated Haar wavelet PURE-LET estimator that is found to compare favorably to the state-of-the-art methods for Poisson intensities estimation, while having a lighter computational burden.

The PURE-LET estimators were finally tested on real multidimensional fluorescence microscopy images. A simple procedure to estimate the various parameters involved in the observation model was proposed and validated experimentally. We also provided some denoising results for 2D+time and 3D datasets, which indicates that large fluorescence microscopy data can be efficiently denoised in a reasonable computation time.

10.2 Perspectives

We believe that the generic "URE-LET" strategy presented in this thesis could be extended along the following directions:

Other noise models We only derived unbiased estimates of the MSE for either an AWGN or a Poisson process degraded by an AWGN. In fact, it is possible to find unbiased estimates of the MSE for a much broader class of noise statistics, such as additive perturbations with arbitrary probability density function (PDF) or PDF from the exponential family. This direction of research has been recently taken by Raphan and Simoncelli in [86], and Eldar in [181].

Other quadratic quality metrics We focused on the minimization of the MSE due to its popularity and its convenient mathematical properties. However, unbiased estimates can be derived for other quadratic quality metrics, such as the weighted MSE (Proposition 1 of Section 2.3). A possible axis of research could consist in designing a perceptual-based weighting matrix.

In the case of a Poisson process with *strictly* positive intensities, the normalized MSE defined as

$$\text{NMSE} = \frac{1}{N}\sum_{n=1}^{N}\frac{(f_n(\mathbf{y}) - x_n)^2}{x_n} \tag{10.1}$$

is sometimes used to evaluate the denoising quality (e.g. in [159]). Using a similar result to the one given in Lemma 4 of Section 6.3, it is also possible to derive an unbiased estimate of the NMSE.

Inverse problems We restricted our investigations to a pure denoising scenario. However, in many applications, the following linear inverse problem (that must be solved for $\mathbf{x}$) is often encountered:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{b} \tag{10.2}$$

where $\mathbf{H}$ is an arbitrary (but deterministic) $L \times N$ ($L \geq N$) distorsion matrix and $\mathbf{b}$ is an additive random perturbation independent of $\mathbf{x}$. $\mathbf{x}$ is estimated from the observed distorted measurements $\mathbf{y}$, as $\hat{\mathbf{x}} = \mathbf{F}(\mathbf{y})$, where $\mathbf{F}$ is a possibly non-linear processing involving several parameters that need to be optimized in the minimum MSE sense. Therefore, a generalized "URE-LET" strategy seems to be appropriate for solving such problem.

When $\mathbf{H} = \mathbf{Id}$, we come back to the denoising problem taken up in this thesis. If $\mathbf{H}$ (or $\mathbf{H}^{\mathrm{T}}\mathbf{H}$) is non-singular, then a SURE-like unbiased MSE estimate can be straightforwardly obtained. However, when $\mathbf{H}$ (or $\mathbf{H}^{\mathrm{T}}\mathbf{H}$) is singular, which might be the case for convolutive distortion, the obtention of an unbiased estimate of the MSE for an arbitrary (non-linear) processing is much more challenging. Several works along this line have been recently undertaken [85, 181–184].

Density estimation We pointed out in Section 2.5 (and 6.5) that the expression of the optimal MMSE estimator involves the PDF of the noisy data $r(\mathbf{y})$. If we assume that a LET estimator constitutes a good approximation of the MMSE estimator, we can then use it to estimate $r(\mathbf{y})$. For instance, consider the AWGN scenario, where the *pointwise* MMSE estimator is given by $\theta_{\mathrm{opt}}(y) = y + \sigma^2 \frac{\mathrm{d}}{\mathrm{d}y} \ln r(y)$, and a LET estimator of the form $\theta_{\mathrm{LET}}(y) = y + \sum_{k=1}^{K} a_k \theta_k(y)$. If one assumes that, for a suitable choice of θ_k, $\theta_{\mathrm{LET}}(y) \simeq \theta_{\mathrm{opt}}(y)$, the PDF of the noisy data could be estimated as:

$$\hat{r}(y) = \mathrm{cst} \cdot \exp\left(\frac{1}{\sigma^2} \sum_{k=1}^{K} a_k \int \theta_k(y)\mathrm{d}y\right)$$

where $\mathrm{cst} = \left(\int_{-\infty}^{\infty} \exp\left(\frac{1}{\sigma^2} \sum_{k=1}^{K} a_k \int \theta_k(y)\mathrm{d}y\right) \mathrm{d}y\right)^{-1}$.

Note that the MMSE estimator could also be used to find out which prior (among a family of parametric densities) is the best representative of the PDF of the noise-free data.

Bibliography

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

[2] A. Eskicioglu and P. Fisher, "Image quality measures and their performance," *Communications, IEEE Transactions on*, vol. 43, no. 12, pp. 2959–2965, Dec 1995.

[3] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70, no. 3, pp. 177 – 200, November 1998.

[4] Z. Wang and A. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81–84, Mar 2002.

[5] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, April 2004.

[6] H. Sheikh and A. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[7] A. Buades, B. Coll, and J. M. Morel, "A Review of Image Denoising Algorithms, with a New One," *Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, vol. 4, no. 2, pp. 490–530, July 2005.

[8] M. Rudin, *Molecular Imaging: Basic Principles and Applications in Biomedical Research*. Imperial College Press, August 2005.

[9] A. Geissler, A. Gartus, T. Foki, A. R. Tahamtan, R. Beisteiner, and M. Barth, "Contrast-to-noise ratio (cnr) as a quality parameter in fmri," *Journal of Magnetic Resonance Imaging*, vol. 25, pp. 1263–1270, 2007.

[10] Z. Wang, H. Sheikh, and A. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, 2002, pp. I–477–I–480 vol.1.

[11] H. Sheikh, A. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: Jpeg2000," *Image Processing, IEEE Transactions on*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

[12] M. C. Motwani, M. C. Gadiya, R. C. Motwani, and F. C. Harris, "Survey of image denoising techniques," in *Proceedings of Global Signal Processing Expo and Conference (GSPx'04)*, Santa Clara, California, September 2004.

[13] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[14] G. J. Burton and I. R. Moorhead, "Color and spatial structure in natural scenes," *Appl. Opt.*, vol. 26, no. 1, pp. 157–170, 1987.

[15] A. van der Schaaf and J. van Hateren, "Modelling the power spectra of natural images: Statistics and information," *Vision Research*, vol. 36, no. 17, pp. 2759 – 2770, 1996.

[16] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications.* Technology Press of the Massachusetts Institute of Technology, 1949.

[17] L. Yaroslavsky, "Linear and rank adaptive filters for image processing," in *Digital Image Processing and Computer Graphics. Theory and Applications*, L. Dimitrov and E. Wenger, Eds., vol. 58. Vienna, Austria: R. Oldenburg, 1991, p. 374.

[18] L. P. Yaroslavsky, "Local adaptive image restoration and enhancement with the use of dft and dct in a running window," in *Wavelet Applications in Signal and Image Processing IV*, M. A. Unser, A. Aldroubi, and A. F. Laine, Eds., vol. 2825, no. 1. SPIE, 1996, pp. 2–13.

[19] J. S. Lee, "Digital Image Enhancement and Noise Filtering by Use of Local Statistics," *IEEE Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 165–168, March 1980.

[20] M. Sedaaghi, "Morphological operators," *IEEE Electronics Letters*, vol. 38, no. 22, October 2002.

[21] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," *Proceedings of the Sixth International Conference on Computer Vision*, pp. 839–846, 1998.

[22] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 60–65 vol. 2, June 2005.

[23] A. Buades, B. Coll, and J. Morel, "Image denoising by non-local averaging," *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 2, pp. 25–28, 18-23, 2005.

[24] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *Image Processing, IEEE Transactions on*, vol. 15, no. 10, pp. 2866–2878, Oct. 2006.

[25] D. V. D. Ville and M. Kocher, "Sure-based non-local means," *Submitted to IEEE Signal Processing Letters*, 2009.

[26] P. Perona and J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, July 1990.

[27] J. Weickert, *Anisotropic Diffusion in Image Processing.* Stuttgart: Teubner-Verlag, 1998.

[28] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[29] T. Chan, S. Osher, and J. Shen, "The digital tv filter and nonlinear denoising," *Image Processing, IEEE Transactions on*, vol. 10, no. 2, pp. 231–241, Feb 2001.

[30] S. Osher, L. I. Rudin, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[31] C. R. Vogel and M. E. Oman, "Fast, Robust Total Variation-Based Reconstruction of Noisy, Blurred Images," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 813–824, June 1998.

[32] M. Nikolova, "Minimizers of Cost-Functions Involving Nonsmooth Data-Fidelity Terms. Application to the Processing of Outliers," *SIAM J. Numer. Anal.*, vol. 40, no. 3, pp. 965–994, 2002.

[33] M. Elad, "Why simple shrinkage is still relevant for redundant representations?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5559–5569, Dec. 2006.

[34] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[35] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, ser. Monographs on Statistics and Applied Probability. New York: Chapman & Hall, 1986.

[36] J. Fessler and I. Hero, A.O., "Penalized maximum-likelihood image reconstruction using space-alternating generalized em algorithms," *Image Processing, IEEE Transactions on*, vol. 4, no. 10, pp. 1417–1429, Oct 1995.

[37] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," *Comm. Pure & Appl. Math.*, vol. 41, pp. 909–996, 1988.

[38] 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, July 1989.

[39] I. Daubechies, "Ten Lectures on Wavelets," *CBMS-NSF Regional Conference series in Applied Mathematics*, vol. 61 of Proc., March 1992.

[40] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transfom," *Computers, IEEE Transactions on*, vol. C-23, no. 1, pp. 90–93, Jan. 1974.

[41] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[42] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Science USA*, vol. 100, pp. 2197–2202, 2003.

[43] R. R. Coifman and D. L. Donoho, "Translation invariant de-noising," in *Lecture Notes in Statistics: Wavelets and Statistics*, vol. 103, Springer Verlag, NewYork, 1995, pp. 125–150.

[44] A. P. Bradley, "Shift-invariance in the discrete wavelet transform," in *Proc. VIIth Digital Image Computing: Techniques and Applications*, O. S. Sun C., Talbot H. and A. T. (Eds.), Eds., Sydney, 10-12 December 2003.

[45] Y. Wan and R. Nowak, "Quasi-circular rotation invariance in image denoising," in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol. 1, 1999, pp. 605–609 vol.1.

[46] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable Multi-scale Transforms," *IEEE Transaction on Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992.

[47] E. Simoncelli and W. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *Image Processing, 1995. Proceedings., International Conference on*, vol. 3, Oct 1995, pp. 444–447 vol.3.

[48] W. Freeman and E. Adelson, "The design and use of steerable filters," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 13, no. 9, pp. 891–906, Sep 1991.

[49] E. Simoncelli and H. Farid, "Steerable wedge filters for local orientation analysis," *Image Processing, IEEE Transactions on*, vol. 5, no. 9, pp. 1377–1382, Sep 1996.

[50] M. Jacob and M. Unser, "Design of steerable filters for feature detection using Canny-like criteria," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1007–1019, August 2004.

[51] P. Vandergheynst and J.-F. Gobbers, "Directional dyadic wavelet transforms: design and algorithms," *Image Processing, IEEE Transactions on*, vol. 11, no. 4, pp. 363–372, Apr 2002.

[52] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The Curvelet Transform for Image Denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, June 2002.

[53] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling & Simulation*, vol. 5, no. 3, pp. 861–899, 2006. [Online]. Available: http://link.aip.org/link/?MMS/5/861/1

[54] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, and P. Dragotti, "Directionlets: anisotropic multidirectional representation with separable filtering," *Image Processing, IEEE Transactions on*, vol. 15, no. 7, pp. 1916–1933, July 2006.

[55] M. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[56] N. G. Kingsbury, "Image Processing with Complex Wavelets," *Phil. Trans. R. Soc. A.*, September 1999.

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

[58] C. Chaux, L. Duval, and J.-C. Pesquet, "Image analysis using a dual-tree m-band wavelet transform," *Image Processing, IEEE Transactions on*, vol. 15, no. 8, pp. 2397–2412, Aug. 2006.

[59] D. Van De Ville and M. Unser, "Complex wavelet bases, steerability, and the Marr-like pyramid," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2063–2080, November 2008.

[60] D. Van De Ville, T. Blu, and M. Unser, "Isotropic polyharmonic B-Splines: Scaling functions and wavelets," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1798–1813, November 2005.

[61] J.-L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*, 2nd ed., ser. Astronomy and Astrophysics Library. Springer, 2006.

[62] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel pca and de-noising in feature spaces," in *Proceedings of the 1998 conference on Advances in neural information processing systems II.* MIT Press, 1999, pp. 536–542.

[63] D. Muresan and T. Parks, "Adaptive principal components and image denoising," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 1, Sept. 2003, pp. I–101–4 vol.1.

[64] T. Sikora and B. Makai, "Shape-adaptive dct for generic coding of video," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 5, no. 1, pp. 59–62, Feb 1995.

[65] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images," *Image Processing, IEEE Transactions on*, vol. 16, no. 5, pp. 1395–1411, May 2007.

[66] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *Image Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 423–438, April 2005.

[67] S. Mallat, "Geometrical grouplets," *Applied and Computational Harmonic Analysis*, vol. 26, no. 2, pp. 161 – 180, 2009.

[68] D. L. Donoho, "Wedgelets: Nearly minimax estimation of edges," *The Annals of Statistics*, vol. 27, no. 3, pp. 859–897, 1999.

[69] R. Willett and R. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 3, pp. 332–350, March 2003.

[70] C. Stein, "Estimation of the Mean of a Multivariate Normal Distribution," *The Annals of Statistics*, vol. 9, pp. 1135–1151, 1981.

[71] 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.

[72] J.-C. Pesquet and D. Leporini, "A New Wavelet Estimator for Image Denoising," *Sixth International Conference on Image Processing and its Applications*, vol. 1, pp. 249–253, July 14-17 1997.

[73] A. Benazza-Benyahia and J.-C. Pesquet, "Building Robust Wavelet Estimators for Multicomponent Images Using Stein's Principle," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1814–1830, November 2005.

[74] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet, "A nonlinear stein-based estimator for multichannel image denoising," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 56, no. 8, pp. 3855–3870, Aug. 2008.

[75] X.-P. Zhang and M. D. Desai, "Adaptive Denoising based on SURE Risk," *IEEE Signal Processing Letters*, vol. 5, no. 10, October 1998.

[76] D. L. Donoho and I. M. Johnstone, "Ideal Spatial Adaptation via Wavelet Shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[77] L. Breiman, "Better Subset Regression Using the Non-Negative Garrote," *Technometrics*, vol. 37, no. 4, pp. 373–384, November 1995.

[78] H.-Y. Gao and A. G. Bruce, "Waveshrink with Firm Shrinkage," *Statistica Sinica*, vol. 7, pp. 855–874, 1997.

[79] ——, "Wavelet Shrinkage Denoising Using the Non-Negative Garrote," *J. Comput. Graph. Stat.*, vol. 7, no. 4, pp. 469–488, 1998.

[80] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive Wavelet Thresholding for Image Denoising and Compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, September 2000.

[81] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, November 2003.

[82] A. Pižurica and W. Philips, "Estimating the Probability of the Presence of a Signal of Interest in Multiresolution Single- and Multiband Image Denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 3, March 2006.

[83] 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, November 2002.

[84] ——, "Bivariate Shrinkage With Local Variance Estimation," *IEEE Signal Processing Letters*, vol. 9, no. 12, December 2002.

[85] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *Preprint*, 2009.

[86] M. Raphan and E. P. Simoncelli, "Learning to be Bayesian without Supervision," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. C. Platt, and T. Hoffman, Eds., vol. 19. Vancouver BC Canada: MIT Press, 4-7 December 2007, pp. 1170–1177.

[87] P. du Bois-Reymond, "Erläuterungen zu den Anfangsgründen der Variationsrechnung," *Mathematische Annalen*, vol. 15, pp. 283 – 314, 1879.

[88] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[89] D. W. Scott and S. R. Sain, "Multidimensional density estimation," in *Data Mining and Data Visualization*, ser. Handbook of Statistics, E. W. C.R. Rao and J. Solka, Eds. Elsevier, 2005, vol. 24, pp. 229 – 261.

[90] G. Nason and B. W. Silverman, *The Stationary Wavelet Transform and some Statistical Applications.* New York: Springer-Verlag, 1995, vol. 103.

[91] N. G. Kingsbury, "Complex Wavelets for Shift Invariant Analysis and Filtering of Signals," *Journal of Applied Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.

[92] F. Luisier, T. Blu, and M. Unser, "A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, March 2007.

[93] F. Luisier and T. Blu, "SURE-LET Multichannel Image Denoising: Interscale Orthonormal Wavelet Thresholding," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 482–492, April 2008.

[94] W. James and C. Stein, "Estimation with Quadratic Loss," *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, vol. 1, pp. 361–379, 1961.

[95] M. Jansen, M. Malfait, and A. Bultheel, "Generalized cross validation for wavelet thresholding," *Signal Processing*, vol. 56, pp. 33–44, January 1997.

[96] S. G. Chang, B. Yu, and M. Vetterli, "Spatially Adaptive Wavelet Thresholding with Context Modeling for Image Denoising," *IEEE Transactions on Image Processing*, vol. 9, no. 9, September 2000.

[97] J. Lee and K. Hoppel, "Noise modeling and estimation of remotely-sensed images," in *Geoscience and Remote Sensing Symposium, 1989. IGARSS'89. 12th Canadian Symposium on Remote Sensing., 1989 International*, vol. 2, Jul 1989, pp. 1005–1008.

[98] P. Meer, J.-M. Jolion, and A. Rosenfeld, "A fast parallel algorithm for blind estimation of noise variance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 2, pp. 216–223, February 1990.

[99] S. I. Olsen, "Noise variance estimation in images," in *Proc. 8th SCIA*, 1993, pp. 25–28.

[100] J. Immerkær, "Fast noise variance estimation," *Comput. Vis. Image Underst.*, vol. 64, no. 2, pp. 300–302, 1996.

[101] S. Beheshti and M. Dahleh, "Noise variance in signal denoising," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 6, April 2003, pp. VI–185–8 vol.6.

[102] J. Han and L. Xu, "A new method for variance estimation of white noise corrupting a signal," in *Instrumentation and Measurement Technology Conference, 2006. IMTC 2006. Proceedings of the IEEE*, April 2006, pp. 1447–1451.

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

[104] P. Vaidyanathan and T. Nguyen, "Eigenfilters: A new approach to least-squares fir filter design and applications including nyquist filters," *Circuits and Systems, IEEE Transactions on*, vol. 34, no. 1, pp. 11–23, Jan 1987.

[105] A. Pižurica, W. Philips, and P. Scheunders, "Wavelet domain denoising of single-band and multiband images adapted to the probability of the presence of features of interest," *Proc. SPIE*, vol. 5914, no. 59141I, September 2005.

[106] P. Scheunders, "Wavelet Thresholding of Multivalued Images," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 475–483, April 2004.

[107] P. Scheunders and J. Driesen, "Least-Squares Interband Denoising of Color and Multispectral Images," *IEEE International Conference on Image Processing*, vol. 2, pp. 985–988, October 2004.

[108] P. Scheunders and S. D. Backer, "Wavelet denoising of multicomponent images, using a gaussian scale mixture model," *18th International Conference on Pattern Recognition*, vol. 3, pp. 754–757, 2006.

[109] N.-X. Lian, V. Zagorodnov, and Y.-P. Tan, "Edge-Preserving Image Denoising via Optimal Color Space Projection," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2575–2587, September 2006.

[110] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian Tree-Structured Image Modeling Using Wavelet-Domain Hidden Markov Models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, July 2001.

[111] T. Blu and F. Luisier, "The SURE-LET Approach to Image Denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2778–2786, November 2007.

[112] F. Luisier and T. Blu, "SURE-LET multichannel image denoising: Undecimated wavelet thresholding," in *Proceedings of the Thirty-Third IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'08)*, Las Vegas NV, USA, March 30-April 4, 2008, pp. 769–772.

[113] J. Gauthier, L. Duval, and J.-C. Pesquet, "Two denoising SURE-LET methods for complex oversampled subband decompositions," in *Proceedings of the 16th European Signal Processing Conference (EUSIPCO'08)*, Lausanne, Switzerland, August 25-29, 2008.

[114] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1540–1554, September 2008.

[115] M. F. Hutchinson, "A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines," *Communications in Statistics - Simulation and Computation*, vol. 19, pp. 433–450, 1990.

[116] C. Bekas, E. Kokiopoulou, and Y. Saad, "An estimator for the diagonal of a matrix," *Appl. Numer. Math.*, vol. 57, no. 11-12, pp. 1214–1229, 2007.

[117] M. Raphan and E. P. Simoncelli, "Optimal denoising in redundant representations," *Image Processing, IEEE Transactions on*, vol. 17, no. 8, pp. 1342–1352, Aug. 2008.

[118] J.-L. Starck, D. L. Donoho, and E. J. Candes, "Very high quality image restoration by combining wavelets and curvelets," in *Wavelets: Applications in Signal and Image Processing IX*, A. F. Laine, M. A. Unser, and A. Aldroubi, Eds., vol. 4478, no. 1. SPIE, 2001, pp. 9–19.

[119] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004.

[120] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

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

[122] A. C. Bovik, *Handbook of Image and Video Processing*, 2nd ed., A. C. Bovik, Ed. San Diego CA, USA: Elsevier Academic Press, June 2005.

[123] M. Ozkan, M. Sezan, and A. Tekalp, "Adaptive motion-compensated filtering of noisy image sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 4, pp. 277–290, Aug 1993.

[124] J. C. Brailean, R. P. Kleihorst, S. Efstratiadis, A. K. Katsaggelos, and R. L. Lagendijk, "Noise Reduction Filters for Dynamic lmage Sequences: A Review," *Proc. of the IEEE*, vol. 83, no. 9, pp. 1272–1292, September 1995.

[125] J. Kim and J. Woods, "Spatio-temporal adaptive 3-d kalman filter for video," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 414–424, Mar 1997.

[126] F. Cocchia, S. Carrato, and G. Ramponi, "Design and real-time implementation of a 3-d rational filter for edge preserving smoothing," *IEEE Trans. Consum. Electron.*, vol. 43, pp. 1291–1300, 1997.

[127] I. W. Selesnick and K. Y. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *SPIE Conf. on Wavelets: Applications in Signal and Image Processing X*, vol. 5207, November 2003, pp. 607–618.

[128] A. Pižurica, V. Zlokolica, and W. Philips, "Noise reduction in video sequences using wavelet-domain and temporal filtering," *SPIE Conf. on Wavelet Applications in Industrial Processing*, vol. 5266, pp. 48–59, February 2004.

[129] V. Zlokolica, A. Pižurica, and W. Philips, "Wavelet-Domain Video Denoising Based on Reliability Measures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 8, pp. 993–1007, August 2006.

[130] E. J. Balster, Y. F. Zheng, and R. L. Ewing, "Combined Spatial and Temporal Domain Wavelet Shrinkage Algorithm for Video Denoising," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 2, pp. 220–230, February 2006.

[131] N. Gupta, M. Swamy, and E. Plotkin, "Wavelet domain-based video noise reduction using temporal discrete cosine transform and hierarchically adapted thresholding," *IET Image Processing*, vol. 1, no. 1, pp. 2–12, January 2007.

[132] S. M. M. Rahman, F. M. Omair Ahmad, and M. N. S. Swamy, "Video Denoising Based on Inter-frame Statistical Modeling of Wavelet Coefficients," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 2, pp. 187–198, February 2007.

[133] G. Varghese and Z. Wang, "Video denoising using a spatiotemporal statistical model of wavelet coefficients," *IEEE Conf. on Acoustics, Speech and Signal Processing ICASSP 2008*, pp. 1257–1260, April 2008.

[134] F. Luisier, T. Blu, and M. Unser, "SURE-LET and block-matching for orthonormal wavelet-domain video denoising," submitted to *IEEE Trans. Circuits Syst. Video Technol.*

[135] J. R. Jain and A. K. Jain, "Displacement Measurement and Its Application in Interframe Image Coding," *IEEE Trans. Commun*, vol. 29, no. 12, pp. 1799–1808, December 1981.

[136] R. Manduchi and G. Mian, "Accuracy analysis for correlation-based image registration algorithms," *IEEE Symp. on Circuits and Systems ISCAS 1993*, pp. 834–837, May 1993.

[137] R. Li, B. Zeng, and M. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, no. 4, pp. 438–442, Aug 1994.

[138] J. Lu and M. Liou, "A simple and efficient search algorithm for block-matching motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 429–433, Apr 1997.

[139] C.-H. Cheung and L.-M. Po, "A novel cross-diamond search algorithm for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 12, pp. 1168–1177, Dec 2002.

[140] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3d transform-domain collaborative filtering," *Proc. of EUSIPCO 2007, Poznań, Poland*, pp. 1257–1260, September 2007.

[141] M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 27–35, Jan. 2009.

[142] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser, "The colored revolution of bioimaging," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 20–31, May 2006.

[143] E. D. Kolaczyk, "Non-parametric estimation of gamma-ray burst intensities using Haar wavelets," *The Astrophysical Journal*, vol. 483, pp. 340–349, 1997.

[144] D. Brillinger, "Some wavelet analyses of point process data," *Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on*, vol. 2, pp. 1087–1091 vol.2, Nov 1997.

[145] K. E. Timmermann and R. D. Nowak, "Multiscale modeling and estimation of poisson processes with application to photon-limited imaging," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 846–862, April 1999.

[146] H. Lu, Y. Kim, and J. M. M. Anderson, "Improved Poisson Intensity Estimation: Denoising Application Using Poisson Data," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1128–1135, August 2004.

[147] E. D. Kolaczyk, "Wavelet shrinkage estimation of certain poisson intensity signals using corrected thresholds," *Statistica Sinica*, vol. 9, pp. 119–135, 1999.

[148] E. D. Kolaczyk and R. D. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Annals of statistics*, vol. 32, no. 2, pp. 500–527, 2004.

[149] E. D. Kolaczyk, "Nonparametric estimation of intensity maps using haar wavelets and poisson noise characteristics," *The Astrophysical Journal*, vol. 534, pp. 490–505, 2000.

[150] ——, "Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds," *Statistica Sinica*, vol. 9, no. 1, pp. 119–135, January 1999.

[151] P. N. Patil and A. T. Wood, "Counting process intensity estimation by orthogonal wavelet methods," *Bernoulli*, vol. 10, no. 1, pp. 1–24, 2004.

[152] S. Sardy, A. Antoniadis, and P. Tseng, "Automatic smoothing with wavelets for a wide class of distributions," *Journal of Computational & Graphical Statistics*, vol. 13, no. 2, pp. 399–423, June 2004.

[153] R. M. Willett and R. D. Nowak, "Multiscale Poisson Intensity and Density Estimation," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3171–3187, September 2007.

[154] D. L. Donoho, "Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data," in *Different Perspectives on Wavelets*, ser. Proceedings of Symposia in Applied Mathematics, I. Daubechies, Ed., vol. 47. Providence, RI, USA: American Mathematical Society, 1993, pp. 173–205.

[155] F. J. Anscombe, "The Transformation of Poisson, Binomial and Negative-Binomial Data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, December 1948.

[156] P. Fryzlewicz and G. P. Nason, "A Haar-Fisz Algorithm for Poisson Intensity Estimation," *Journal of Computational and Graphical Statistics*, vol. 13, no. 3, pp. 621–638, September 2004.

[157] P. Besbeas, I. D. Feis, and T. Sapatinas, "A Comparative Simulation Study of Wavelet Shrinkage Estimators for Poisson Counts," *International Statistical Review*, vol. 72, no. 2, pp. 209–237, August 2004.

[158] M. Jansen, "Multiscale poisson data smoothing," *Journal of the Royal Statistical Society B*, vol. 68, no. 1, pp. 27–48, 2006.

[159] B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, ridgelets, and curvelets for poisson noise removal," *Image Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1093–1108, July 2008.

[160] R. D. Nowak, "Wavelet-Based Rician Noise Removal for Magnetic Resonance Imaging," *IEEE Transactions on Image Processing*, vol. 8, no. 10, October 1999.

[161] A. Antoniadis and T. Sapatinas, "Wavelet shrinkage for natural exponential families with quadratic variance functions," *Biometrika*, vol. 88, no. 3, pp. 805–820, 2001.

[162] H. M. Hudson, "A natural identity for exponential families with applications in multiparameter estimation," *The Annals of Statistics*, vol. 6, no. 3, pp. 473–484, May 1978.

[163] K.-W. Tsui and S. J. Press, "Simultaneous Estimation of Several Poisson Parameters Under K-Normalized Squared Error Loss," *The Annals of Statistics*, vol. 10, no. 1, pp. 93–100, March 1982.

[164] S. Ramani, C. Vonesch, and M. Unser, "Deconvolution of 3D fluorescence micrographs with automatic risk minimization," in *Proceedings of the Fifth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08)*, Paris, France, May 14-17, 2008, pp. 732–735.

[165] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of poisson-corrupted images," *Signal Processing*, July 2009.

[166] K. Hirakawa, F. Baqai, and P. J. Wolfe, "Wavelet-based Poisson rate estimation using the Skellam distributions," in *Proceedings of the IS&T/SPIE 20th Annual Symposium on Electronic Imaging Science and Technology*, vol. 7246, 2009.

[167] J. G. Skellam, "The Frequency Distribution of the Difference Between Two Poisson Variates Belonging to Different Populations," *Journal of the Royal Statistical Society*, vol. 109, no. 3, p. 296, 1946.

[168] F. Murtagh, J.-L. Starck, and A. Bijaoui, "Image Restoration with Noise Suppression using a Multiresolution support," *Astron. Astrophys. Suppl. Ser.*, vol. 112, pp. 179–189, 1995.

[169] B. Zhang, J. Fadili, J.-L. Starck, and J.-C. Olivo-Marin, "Multiscale variance-stabilizing transform for mixed-poisson-gaussian processes and its applications in bioimaging," *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 6, pp. VI –233–VI –236, 16 2007-Oct. 19 2007.

[170] M. Yuan, "Automatic smoothing for poisson regression," *Communications in statistics. Theory and methods*, vol. 34, no. 3, pp. 603–617, 2005.

[171] I. Konstantinidis, A. Santamaria-Pang, and I. Kakadiaris, "Frames-based denoising in 3d confocal microscopy imaging," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2005, pp. 290–293.

[172] G. Bernad, L. Blanc-Feraud, and J. Zerubia, "A restoration method for confocal microscopy using complex wavelet transform," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 2, 18-23, 2005, pp. 621–624.

[173] J. Boulanger, C. Kervrann, and P. Bouthemy, *Medical Image Computing and Computer-Assisted Intervention*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2005, ch. Adaptive Spatio-Temporal Restoration for 4D Fluorescence Microscopic Imaging, pp. 893–901.

[174] J. Boulanger, J.-B. Sibarita, C. Kervrann, and P. Bouthemy, "Non-parametric regression for patch-based fluorescence microscopy image sequence denoising," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, May 2008, pp. 748–751.

[175] J. Pawley, *Three-Dimensional Confocal Microscopy: Volume Investigation of Biological Specimens*. Academic Press, 1994, ch. Sources of noise in three-dimensional microscopical data sets, pp. 47–94.

[176] L. Takács, "Über die wahrscheinlichkeitstheoretische Behandlung der Anodenstromschwankungen von Elektronenröhren," *Acta physica Academiae scientiarum hungaricae*, vol. 7, pp. 25–50, 1957.

[177] H. H. Barrett and K. J. Myers, *Foundations of image science*. Wiley, 2003.

[178] A. G. Wright, "The statistics of multi-photoelectron pulse-height distributions," *Nuclear Instruments and Methods in Physics Research A*, vol. 579, pp. 967–972, 2007.

[179] D. L. Snyder and A. M. Hammoud, "Image Recovery from Data Acquired with a Charge-Coupled-Device Camera," *Journal of Optical Society of America A*, vol. 10, no. 5, pp. 1014–1024, 1993.

[180] J. Boulanger, "Estimation non-paramétrique et contributions à l'analyse de séquences d'images. modélisation, simulation et estimation du trafic intracellulaire dans des séquences de vidéo-microscopie," Ph.D. dissertation, Université de Rennes I, 2007.

[181] Y. Eldar, "Generalized SURE for Exponential Families: Applications to Regularization," *Signal Processing, IEEE Transactions on*, vol. 57, no. 2, pp. 471–481, Feb. 2009.

[182] C. Vonesch, S. Ramani, and M. Unser, "Recursive risk estimation for nonlinear image deconvolution with a wavelet-domain sparsity constraint," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, Oct. 2008, pp. 665–668.

[183] R. Giryes, M. Elad, and Y. Eldar, "Automatic parameter setting for iterative shrinkage methods," in *Electrical and Electronics Engineers in Israel, 2008. IEEEI 2008. IEEE 25th Convention of*, Dec. 2008, pp. 820–824.

[184] M. Ting, R. Raich, and A. Hero, "Sparse image reconstruction for molecular imaging," *Image Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1215–1227, June 2009.

Acknowledgement

During the few years I spent within the Biomedical Imaging Group (BIG) at EPFL, I had the pleasure to benefit from the personal and professional qualities of many former and current BIG members. I take this opportunity to warmly thank all of them:

- Michael, the BIG director and incidentally, my thesis advisor, for having made me very welcome in his lab and for all his support. In particular, he always reminds me of the biomedical application behind this work, which, at first sight, could have looked like a "denoise *Lena* toy problem".

- Thierry, former BIG project leader, now professor at the Chinese University of Hong Kong (CUHK) and incidentally, my thesis co-advisor, for having shared a part of his unbounded (if there were some bounds, I could not find them anyway) scientific knowledge. He showed me that the simplest solutions are not the easiest to find, but are often the most beautiful. Throughout this thesis, I could benefit from his inspired intuitions and his constant enthusiasm. I would also like to thank him for having welcome me in his lab at CUHK and for the great moments we spent during this two-month stay in Hong Kong.

- Daniel, the BIG head of software development, for his constant availability, for his help in solving all soft- and hardware issues, as well as for his precious support in *java* programming.

- Philipe, the BIG first research assistant, for ensuring the smooth running of the whole lab, and for having reviewed the abstract of this thesis.

- Dimitri, former BIG project leader, now professor at the University of Geneva, for having accepted me in his office during my master thesis at BIG, for the fruitful discussions we had, and for his communicative good mood.

- Manuelle, the BIG secretary, for her help in carrying out all my administrative tasks.

- Sathish and Jean-Charles, my former and my current BIG office mates, for all the great moments we spend together in our "small" office, talking about work, life and entertainment.

- Cédric, former BIG member, for the many fruitful discussions we had about Poisson noise reduction, as well as for the great games of tennis we played.

- All current and former BIG members, for all the interactions we had, either professional or personal: (in no order of preference) Chandra, Ilker, Hagai, Aurélien, Kunal, Ricard, Matthieu, Djano, François, Pouya, Raquel, Kai, Simona, Katarina, Ildar.

I am also very grateful to the members of my thesis jury Prof. Christian Depeursinge, Prof. Jean-Christophe Pesquet, Prof. Jalal Fadili and Prof. Pierre Vanderghyenst for having accepted to read and comment the present work.

Last but not least, I would like to thank all my family: my father René, my mother Anne, my brother Mathieu and my sister Agathe for everything. Nothing could be possible without you.

Florian Luisier

Swiss Federal Institute of Technology (EPFL)
Biomedical Imaging Group
Station 17
CH-1015 Lausanne VD
Switzerland

Phone: +41 (21) 693 51 37
Email: florian.luisier@epfl.ch
URL: http://people.epfl.ch/florian.luisier

PERSONAL DATA

Date of Birth:	September 29, 1981
Citizenship:	Swiss
Marital Status:	Single

RESEARCH INTERESTS

Image processing, multiresolution representations, risk estimation techniques, statistical signal processing, restoration of multidimensional biomedical data.

EDUCATION

Swiss Federal Institute of Technology, Lausanne, Switzerland

- 2005-2009 — **Ph. D.** within the *Biomedical Imaging Group*, graduation in 2009.
  Thesis: "The SURE-LET Approach to Image Denoising"
  Advisors: Prof. Thierry Blu and Prof. Michael Unser

- 2000-2005 — **M. Sc.** in *Microengineering*, graduation in 2005.
  Thesis: "*Thresholding Strategies for Image Denoising with an Application to fMRI Statistical Analysis*"
  Advisors: Prof. Thierry Blu and Prof. Michael Unser

Collège de l'Abbaye, Saint-Maurice, Switzerland

- 1995-2000 — **Maturité fédérale, type C (scientific)**, graduation in 2000.
  Secondary school examination giving university entrance qualification.
  Swiss equivalent to US high-school diploma.

TEACHING EXPERIENCE

Swiss Federal Institute of Technology, Lausanne, Switzerland

- 2005-2009 **Teaching Assistant** for the Master course "*Image Processing I+II*". **Supervisor** of several semester projects carried out by EPFL Master students.

PROFESSIONAL ACTIVITIES

Membership

- Graduate Student Member of the *Institute of Electrical and Electronics Engineers* (IEEE).
- Member of the *Swiss Society for Biomedical Engineering* (SSBE).

Review

- Reviewer for various international scientific journals, including *IEEE Transactions on Image Processing* and *IEEE Transactions on Signal Processing*.
- Reviewer for the European Signal Processing Conference (EUSIPCO), 2008.

PUBLICATIONS

Journal Articles

1. F. Luisier, C. Vonesch, T. Blu, M. Unser, "Fast Interscale Wavelet Denoising of Poisson-corrupted Images", *Signal Processing*, vol. 90, no. 2, pp. 415-427, February 2010.
2. F. Luisier, T. Blu, "SURE-LET Multichannel Image Denoising: Interscale Orthonormal Wavelet Thresholding," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 482-492, April 2008.
3. T. Blu, F. Luisier, "The SURE-LET Approach to Image Denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2778-2786, November 2007.
4. F. Luisier, T. Blu, M. Unser, "A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593-606, March 2007.

5. Y. Lopez de Meneses, P. Roduit, F. Luisier, J. Jacot, "Trajectory Analysis for Sport and Video Surveillance," *Electronic Letters on Computer Vision and Image Analysis*, vol. 5, no. 3, pp. 148-156, August 2005.

Conference Articles

1. F. Luisier, C. Vonesch, T. Blu, M. Unser, "Fast Haar-Wavelet Denoising of Multidimensional Fluorescence Microscopy Data", *Proceedings of the Sixth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'09)*, Boston MA, USA, June 28-July 1, 2009, pp. 310-313.

2. S. Delpretti, F. Luisier, S. Ramani, T. Blu, M. Unser, "Multiframe SURE-LET Denoising of Timelapse Fluorescence Microscopy Images," *Proceedings of the Fifth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (IS-BI'08)*, Paris, France, May 14-17, 2008, pp. 149-152.

3. F. Luisier, T. Blu, "SURE-LET Multichannel Image Denoising: Undecimated Wavelet Thresholding," *Proceedings of the Thirty-Third IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'08)*, Las Vegas NV, USA, March 30-April 4, 2008, pp. 769-772.

4. F. Luisier, T. Blu, "SURE-LET Interscale-Intercolor Wavelet Thresholding for Color Image Denoising," *Proceedings of the SPIE Optics and Photonics 2007 Conference on Mathematical Methods: Wavelet XII*, San Diego CA, USA, August 26-29, 2007, vol. 6701, pp. 67011H-1/67011H-10.

5. F. Luisier, T. Blu, "Image Denoising by Pointwise Thresholding of the Undecimated Wavelet Coefficients: A Global SURE Optimum," *Proceedings of the Thirty-Second IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'07)*, Honolulu HI, USA, April 15-20, 2007, pp. I-593-I-596.

6. F. Luisier, T. Blu, M. Unser, "SURE-Based Wavelet Thresholding Integrating Inter-Scale Dependencies," *Proceedings of the 2006 IEEE International Conference on Image Processing (ICIP'06)*, Atlanta GA, USA, October 8-11, 2006, pp. 1457-1460.

7. F. Luisier, T. Blu, B. Forster, M. Unser, "Which Wavelet Bases Are the Best for Image Denoising ?," *Proceedings of the SPIE Optics and Photonics 2005 Conference on Mathematical Methods: Wavelet XI*, San Diego CA, USA, July 31-August 3, 2005, vol. 5914, pp. 59140E-1/59140E-12.

AWARDS

- **2009 Young Author Best Paper Award from the IEEE Signal Processing Society** for the paper co-authored with T. Blu and M. Unser, "A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding", IEEE Trans. Image Processing, Vol. 16, no. 3, pp. 593 - 606, March 2007.
- **Student travel grant** to attend the *Thirty-Second IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'07).*
- **Finalist of the Student Paper Contest** of the *Thirty-Second IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'07).*