Multiscale Statistical Image Models and Bayesian Methods

Aleksandra Pižurica and Wilfried Philips

Ghent University, Dept. Telecommunications and Information Processing,
Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium

ABSTRACT

Multiscale statistical signal and image models resulted in major advances in many signal processing disciplines. This paper focuses on Bayesian image denoising. We discuss two important problems in specifying priors for image wavelet coefficients. The first problem is the characterization of the marginal subband statistics. Different existing models include highly kurtotic heavy-tailed distributions, Gaussian scale mixture models and weighted sums of two different distributions. We discuss the choice of a particular prior and give some new insights in this problem. The second problem that we address is statistical modelling of inter- and intrascale dependencies between image wavelet coefficients. Here we discuss the use of Hidden Markov Tree models, which are efficient in capturing inter-scale dependencies, as well as the use of Markov Random Field models, which are more efficient when it comes to spatial (intrascale) correlations. Apart from these relatively complex models, we review within a new unifying framework a class of low-complexity locally adaptive methods, which encounter the coefficient dependencies via local spatial activity indicators.

Keywords: Wavelets, Bayesian estimation, Markov Random Field models, Hidden Markov Tree models.

1. INTRODUCTION

Statistical modelling of image features at multiple resolution scales is a topic of tremendous interest for numerous disciplines including image restoration, image analysis and segmentation, data fusion... A number of comprehensive publications on this subject include tutorials [1], special issues [2,3] and books [4]. Multiscale stochastic signal and image models are usually linked to wavelet representation [5–8], which provides a natural framework for multiresolution analysis. Here we focus on wavelet domain image denoising where different stochastic models for wavelet coefficients are used within a Bayesian estimation approach.

1.1. Bayesian wavelet shrinkage

Noise reduction in the wavelet domain usually results from *wavelet shrinkage*: ideally, the coefficients that contain primarily noise should be reduced to negligible values while the ones containing a "significant" noise-free component should be reduced less. A common shrinkage approach is thresholding [9, 10], where the coefficients with magnitudes below a certain threshold are treated as "non significant" and are set to zero. The remaining, "significant" coefficients are kept unmodified (hard-thresholding) or they are reduced in magnitude (soft-thresholding).

Shrinkage estimators can also result from a *Bayesian approach* which imposes a prior distribution [11–13] of noise-free data. The simplest Bayesian methods assume statistically independent data and rely on marginal statistics only [14–19]. Others encounter prior knowledge about inter- and/or intrascale dependencies among the coefficients as well, by using bivariate [20] or joint [21] statistics, by employing Hidden Markov Tree (HMT) models [22–26] or Markov Random Field (MRF) models [27–30], or alternatively, by using some local (context) measurements calculated from a surrounding of each coefficient [31–37]. Another possible categorization is according to the optimization criterion employed, i.e., according to the adopted estimation rule. Here we recognize at least three classes:

(i) *methods that optimize the threshold selection for hard- and soft-thresholding* [11, 12, 14, 15]. For example, the soft thresholding method of [14], employs a threshold that is optimal in terms of mean squared error under marginal subband statistics of natural images. Its spatially adaptive extension is in [31].

Send correspondence to Aleksandra.Pizurica@telin.ugent.be.

$w_{j,l}^D$	wavelet coefficient: scale 2^j, spatial position l and orientation D
w_l	wavelet coefficient at the position l in a given subband
y_l	noise-free coefficient value
$\hat{y}_l$	an estimate of y_l
m_l	significance measure for w_l
x_l	hidden variable, binary label for w_l
$\mathbf{w}$, $\mathbf{m}$, $\mathbf{x}$	vectors: detail image, significance map and mask, resp.
W_l, M_l, X_l	random variables
$\mathbf{W}$, $\mathbf{M}$, $\mathbf{X}$	random vectors
∂l	neighborhood of the pixel l
$\mathbf{X}_{\mathcal{A}}$	$\{\mathrm{x}_l : l \in \mathcal{A}\}$
$P(A = a\|B = b)$	conditional probability of a given b
$f_Y(y), f(y)$	probability density function of y
$f_{W\|X}(w\|x)$	conditional probability density function of w given x

Table 1. Nomenclature.

(ii) *estimators that result from minimizing a Bayesian risk*, typically under a quadratic cost function (minimum mean squared error - MMSE estimation [16–18,20,38]) or under a delta cost function (maximum a posteriori - MAP estimation [19]). Spatially adaptive extensions of such estimators are, e.g., [22–24, 32–36].

(iii) *methods that multiply each wavelet coefficient with the probability that it contains a significant noise-free component* (given a set of measurements calculated from the empirical coefficients) [27–30, 37]. Like hard- or soft-thresholding functions, such a shrinkage rule seems rather ad-hoc but is also intuitively appealing and effective in practice. A recent more extensive analysis of this type of estimators is in [39].

1.2. Notation and Noise Model

In the sequel, we use the following notation: a wavelet coefficient $w_{j,l}^D$ represents the bandpass content of an image at resolution scale 2^j, spatial position l and orientation D. Whenever there can be no confusion, we omit the indices that denote the scale and the orientation. Random variables are denoted by capital letters and their realizations by the corresponding small letters. Boldface letters are used for vectors. A detail image (wavelet subband) is represented as $\mathbf{w} = \{w_1, ..., w_n\}$, where the set of indices $L = \{1, ..., n\}$ is a set of pixels on a regular rectangular lattice. We shall often assign a significance measure m_l and a binary label x_l to each wavelet coefficient w_l. For example, the label value $x_l = 0$ denotes that w_l represents mainly noise, and the value $x_l = 1$ denotes that w_l is a "significant" coefficient. A set of these labels $\mathbf{x} = \{x_1, ..., x_n\}$ is called *mask*, while the set of significance measures $\mathbf{m} = \{m_1, ..., m_n\}$ is called *significance map*. The nomenclature is in Table 1. We often abbreviate the probability density function by *density*.

Unless otherwise stated, we assume *additive white Gaussian noise*: $w_l = y_l + n_l$, where y_l is the noise-free coefficient component and n_l are independent, identically distributed zero mean normal random variables $N(0, \sigma^2)$. An *orthogonal* wavelet transform maps the white noise in the input image into a white noise in the wavelet domain. In this case, the noise standard deviation in each detail image is equal to the standard deviation of the input noise σ. In a *non-decimated* representation the noise contributions n_l are not independent and the noise standard deviation σ_j^D depends on the resolution level j and on the subband orientation D.

Only in Sections 4.2 and 4.3 we illustrate the suppression of other than Gaussian noise types.

1.3. Contents and structure

In Section 2 we discuss different marginal priors for noise free subband data. Denoising approaches based on HMT and MRF models are discussed in Section 3. Here we outline the main ideas, as well as conceptual differences and similarities between these methods. In Section 4, we review within a new unifying framework a low-complexity locally adaptive approach, which encounters the coefficient dependencies via local spatial activity indicators. The conclusions are in Section 5.

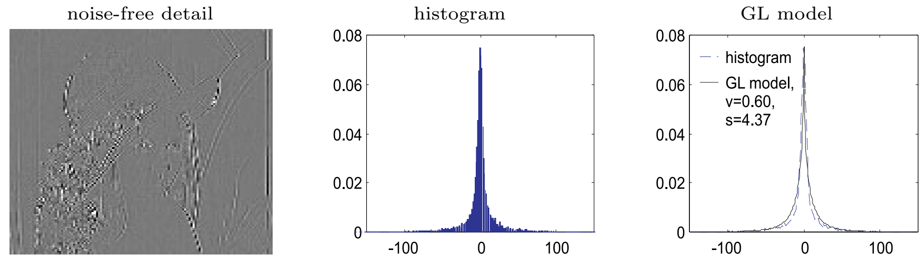


Figure 1. An illustration of the generalized Laplacian (GL) prior for noise-free wavelet coefficients.

2. MARGINAL STATISTICS OF IMAGE WAVELET COEFFICIENTS

In a wavelet decomposition of a noise-free image many wavelet coefficients come from relatively smooth regions and are thus quite small, while others corresponding to edges can be very large. Hence, as discussed by many authors (e.g., [6, 18, 19, 21, 22, 32]) the distribution of noise-free wavelet coefficients in each subband is sharply peaked at zero and heavy tailed.

2.1. Heavy tailed distributions

A common marginal prior for noise-free subband data is *generalized Laplacian* (also called *generalized Gaussian* distribution) [14, 18, 40]

$$f(y) = \frac{\nu}{2s\Gamma(\frac{1}{\nu})} \exp(-|y/s|^{\nu}), \quad s, \nu > 0 \tag{1}$$

where $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ is the Gamma function. For natural images, the shape parameter is typically $\nu \in [0, 1]$. The variance and the courtosis of a generalized Laplacian signal are [18] $\sigma_y^2 = s^2\Gamma(\frac{3}{\nu})/\Gamma(\frac{1}{\nu})$, and $\kappa_y = \Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})/\Gamma^2(\frac{3}{\nu})$, respectively. The model parameters s and ν are accurately estimated from a signal corrupted by additive white Gaussian noise [18].

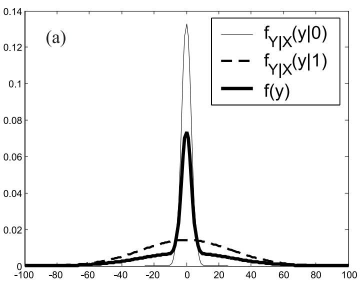
A special case in the family (1), with $\nu = 1$ (the so-called *Laplacian* or *double exponential*) is often used because of its analytical tractability [19, 38] since it usually does not produce a noticeable degradation in performance. The MAP estimation under the Laplacian prior yields a soft-thresholding function with the threshold $\sqrt{2}\sigma^2/\sigma_y$ [19]. The MMSE estimate is derived in [38]. Other heavy tailed distributions of wavelet coefficients have been proposed for specific types of images, like, e.g., the *Pearson distributions* for SAR images in [41], and the *α-stable distributions* for medical ultrasound images in [42].

2.2. Mixture priors

As compared to generalized Laplacian model, mixture priors [13, 15, 16, 32, 35] often yield a reduced computation complexity of a Bayesian estimator. Moreover, as we show next, mixture priors also offer an elegant way for adapting a Bayesian estimator to the surrounding of each coefficient.

2.2.1. Gaussian scale mixtures

A Gaussian scale mixture prior [32–35] models each coefficient as the product of two independent random variables: $y = \sqrt{z}u$, where z is a positive scalar and u is an element of a Gaussian random field. The multiplier z is usually a function of the surrounding coefficient values (like the local variance of the coefficients within the same scale [32] or a more complex function of the neighboring coefficients within the same and adjacent scales [34, 35]). The MMSE estimate with such priors takes the form of a locally adaptive Wiener-like estimator.



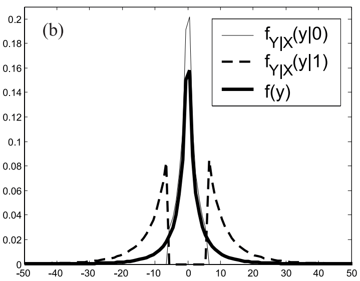


Figure 2. (a) A mixture of two normals: $f_{Y|X}(y|0) = N(0, \sigma_0^2)$, $f_{Y|X}(y|1) = N(0, \sigma_1^2)$. (b) A Laplacian mixture: $f_{Y|X}(y|0) \propto e^{-\lambda|y|}$ for $|y| \leq T$ otherwise zero; $f_{Y|X}(y|1) \propto e^{-\lambda|y|}$ for $|y| > T$ otherwise zero;

2.2.2. Weighted mixtures of two distributions

Another common class of mixture priors, are superpositions of two distributions, where one distribution models the statistics of "significant" ("high energy" or "important") coefficients and the other distribution models the statistics of "non-significant" coefficients and where the mixing parameter is a Bernoulli random variable [12, 13, 15–17, 22–24]. Within this framework, common models are the *mixture of two normals* [16, 22–24] and a mixture of a normal distribution and a point mass at zero [15, 17]. A systematic overview of these and related models is in [12]. The marginal prior of [38] is a mixture of a point mass at zero and the Laplacian distribution. In a sense, a generalization of this prior is in [30], where the distribution of "significant" coefficients is described by the tails of a Laplacian and the distribution of "non significant" coefficients by the central (low-magnitude) part of the same distribution. We can write a unifying form for the above mixture priors as:

$$f(y) = P(X = 0) f_{Y|X}(y|0) + P(X = 1) f_{Y|X}(y|1), \tag{2}$$

where X is a Bernoulli random variable with $P(X = 1) = p = 1 - P(X = 0)$, and where $f_{Y|X}(y|0)$ and $f_{Y|X}(y|1)$ are the densities of "non significant" and "significant" noise-free coefficients, respectively. In some approaches, $P(X = 1)$ is estimated *per subband* [16], while it is in others estimated *adaptively for each coefficient* using e.g., HMT modelling framework [22–24] (Section 3.1), using MRF modelling framework [27–30] (Section 3.2) or by conditioning the probability of signal presence on a local spatial activity indicator [37] (Section 4).

Under the prior (2), the minimum mean squared error estimate of the noise-free coefficient value is

$$E(y|w) = P(X = 0|w) E(y|w, X = 0) + P(X = 1|w) E(y|w, X = 1). \tag{3}$$

Using the Bayes' rule one can show that $P(X = 1|y) = \mu\xi/(1 + \mu\xi)$, where $\mu = P(X = 1)/P(X = 0)$ is the prior ratio and $\xi = f_{Y|X}(y|1)/f_{Y|X}(y|0)$ is the likelihood ratio.

2.3. Comments on the choice of a marginal prior

The choice of a marginal prior is a crucial step for designing a Bayesian denoising method. Even methods that encounter inter- and intrascale dependencies as well, usually build on a given marginal prior for subband data.

The mixture of two normals is attractive due to its analytical tractability. MMSE estimates based on this prior are of simple and elegant form. In particular, if we denote $f_{Y|X}(y|0) = N(0, \sigma_0^2)$ and $f_{Y|X}(y|1) = N(0, \sigma_1^2)$, the conditional means are $E(y|w, H_0) = \sigma_0^2/(\sigma_0^2 + \sigma_1^2)w$ and $E(y|w, H_1) = \sigma_1^2/(\sigma_0^2 + \sigma_1^2)w$. However, this model is not truly heavy-tailed. Since it involves three parameters (two mixing variances and the probability $p = P(X = 1)$) the parameter estimation can present a substantial problem unless they are fixed per subband (like in [16]).

The mixture model from Fig. 2(b) has a natural interpretation: "significant" are the noise-free coefficient magnitudes above a certain threshold while the others are "non significant". Their statistical distributions follow a realistic, Laplacian model. Following the optimum coefficient selection principle [7, 28], the threshold T that defines a significant magnitude should equal the noise standard deviation $T = \sigma$. This leaves only one parameter of the mixture prior. It was shown in [39] that if this parameter is fixed per subband and equal to

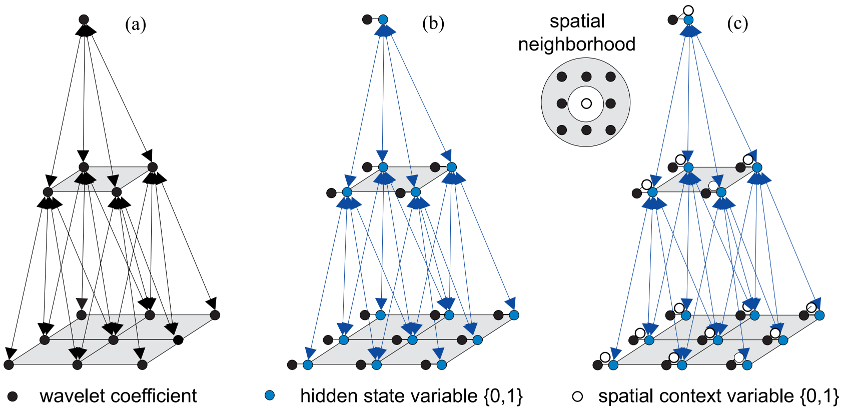


Figure 3. A schematic representation of (a) the multiscale stochastic process on a quadtree used in [44]; (b) the HMT model of [22] and (c) the Local contextual hidden Markov model of [24].

$p = \exp(-\sigma/s)/(1-\exp(-\sigma/s))$, the mixture prior reduces to the Laplacian. The main interest is in estimating the probability $P(X=1)$ adaptively for each coefficient, rather than fixing it per subband. In such a context, the Laplacian mixture prior was used in a MRF based method [30] (see Fig. 5), but not within MMSE estimation. The latter is possible too, and the required conditional means $E(y|w,H_0)$ and $E(y|w,H_1)$ are derived in [39].

Using Gaussian scale mixtures, one can obtain a great variety of truly heavy-tailed distributions [35]. Such models also offer an elegant framework for constructing spatially adaptive estimators that range from low-compexity ones [32] to highly sophisticated ones like [35].

As a final remark, it is interesting to note that the use of mixtures of two distributions (2) seems particularly attractive for applications where in addition to the noisy data some other sources can be used to locate the image edges (like in case of hyperspectral data and other multivalued images). In such applications the probability p that a coefficient at a given spatial position represents a significant edge can be based on data fusion.

3. MODELLING INTER- AND INTRASCALE DEPENDENCIES

A theoretic study of inter- and intrascale dependencies among the wavelet coefficients is in [43]. The Hidden Markov Tree (HMT) models are extensively used in recent wavelet denoising literature, e.g., [22–26]. Use of Markov Random Field (MRF) models for spatial clustering of the coefficients [27–30] has been considerably less studied. This Section outlines the main ideas, differences and similarities between these approaches.

3.1. Hidden Markov Tree (HMT) models

Interscale ("parent - child") dependencies among the wavelet coefficients in a decimated wavelet representation are naturally modelled on a *quadtree* structure. Due to downsampling, each coefficient at the scale 2^j corresponds to four coefficients at the next finer scale 2^{j-1}. Multiscale stochastic processes on quadtrees are studied, e.g., in [44–46], where the wavelet coefficients are modeled using Markov relationships of the type "parent - child" on a quadtree (see Fig. 3(a)). Hidden Markov Tree (HMT) models [22–25] establish similar relationships among the *hidden state variables* rather than among the coefficients themselves (see Fig. 3(b)).

HMT models of [22–24] build on the marginal prior of [16], which is a mixture of two normals (Section 2.2.2). With each wavelet coefficient w_l a hidden state random variable X_l is associated, where $x_l \in \{0,1\}$ and where $x_l = 1$ denotes that w_l contains a large noise-free component, while $x_l = 0$ denotes the opposite. This hidden variable describes the random choice of which mixture component is used for the particular wavelet coefficient. The prior model is a locally adaptive version of (2): $f(y_l) = p_l^0 \mathcal{N}(0,\sigma_{0,l}^2) + p_l^1 \mathcal{N}(0,\sigma_{1,l}^2)$, with $p_l^1 = P(X_l = 1)$, and $p_l^0 = 1 - p_l^1$. In addition, each parent-child state-to-state link has a corresponding state transition matrix

$$A_l = \begin{bmatrix} p_l^{0\to 0} & p_l^{0\to 1} \\ p_l^{1\to 0} & p_l^{1\to 1} \end{bmatrix} \tag{4}$$

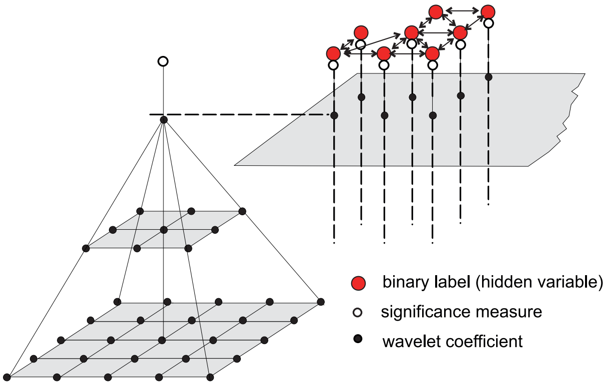


Figure 4. Interactions among the attached (hidden) variables in a MRF based approach. A significance measure that is attached with each wavelet coefficient can be computed from several scales.

with $p_l^{0\to1} = 1 - p_l^{0\to0}$ and $p_l^{1\to0} = 1 - p_l^{1\to1}$. The parameters $p_l^{0\to0}$ and $p_l^{1\to1}$ are the *persistency probabilities*, while $p_l^{0\to1}$ and $p_l^{1\to0}$ are called the *novelty probabilities*, for they express the probability that the state values will change from one scale to the next [23]. The HMT model is specified in terms of (1) the mixture variances $\sigma_{l,0}^2$ and $\sigma_{l,1}^2$; (2) the state transition matrices A_l and (3) the probability of a large state at the root node. These parameters are grouped in a vector $\mathbf{\Theta}$. The conditional mean of y_l given the noisy value w_l and given the parameter vector $\mathbf{\Theta}$ is

$$\hat{y}_l = E(y_l|w_l, \mathbf{\Theta}) = \left(P(X_l = 0|w_l, \mathbf{\Theta}) \frac{\sigma_{l,0}^2}{\sigma^2 + \sigma_{l,0}^2} + P(X_l = 1|w_l, \mathbf{\Theta}) \frac{\sigma_{l,1}^2}{\sigma^2 + \sigma_{l,1}^2} \right) w_l \tag{5}$$

where σ is the noise standard deviation. The required probabilities are estimated by "upward-downward" algorithms through the tree, and using model training procedures detailed in [22]. Commonly mentioned problems are: (1) *a large number of unknown parameters* (implies simplifications such as parameter invariance within the scale) (2) *convergence can be relatively slow* [23] and (3) *lack of spatial adaptation* - the links in the quadtree from Fig. 3(b) do not capture the *intrascale* dependencies. In this respect, a local contextual HMT model of [24] is an improvement: an additional hidden state is attached to each coefficient; this additional hidden variable is a function of the surrounding wavelet coefficients, as illustrated in Fig. 3(c). The actual "interactive communication" between the state variables is still in the vertical direction only and not within the scale.

3.2. Markov Random Field (MRF) models for spatial clustering

In [27–30] a methodology is developed for image denoising based on Markov Random Field models for spatial clustering of image wavelet coefficients. In these approaches, a bi-level MRF model encodes "geometry" of detail images by giving preference to spatially connected configurations of large wavelet coefficients.

To make a parallel with the HMT models, here a binary hidden variable x_l is also attached with each coefficient w_l, where $x_l = 1$ again denotes that w_l contains a significant noise-free component and $x_l = 0$ denotes the opposite. In contrast to HMT models, the interactions among the hidden variables are now *horizontal*, i.e., within the scale. In particular, the vector of binary labels $\mathbf{x} = [x_1, ..., x_n]$ for all the coefficient within a given detail image is called a *mask* and each possible mask is assumed to be a realization of a Markov Random Field $\mathbf{X}$. In a Markov Random Field the probability of a pixel label, given all other pixel labels in the image, reduces to a function of neighboring* labels only [47]. A set of pixels, which are all neighbors of one another is called a *clique*†. The joint probability $P(\mathbf{X} = \mathbf{x})$ of a MRF is a special case of the Gibbs distribution $\exp(-H(\mathbf{x})/T)/Z$,

*Most often used are the so-called first-order neighborhood (four nearest pixels) and the second-order neighborhood (eight nearest pixels).

†For example, for the first-order neighborhood cliques consist of one or two pixels, and for the second order neighborhood cliques consist of up to four pixels.

with partition constant Z and temperature T, where the energy $H(\mathbf{x})$ can be decomposed into contributions of *clique potentials* $V_C(\mathbf{x})$ over all possible cliques:

$$P(\mathbf{X}=\mathbf{x}) = \frac{1}{Z}\exp\Big(-\frac{1}{T}\sum_{C\in\mathcal{C}} V_C(\mathbf{x})\Big). \tag{6}$$

The clique potential $V_C(\mathbf{x})$ is a function of *only* those labels x_l, for which $l \in C$. One defines the appropriate clique potential functions to give preference to certain local spatial dependencies, e.g., to assign higher prior probability to edge continuity. Now we can summarize the essence of denoising methods [27–30] as follows.

1. Assign to each detail image (i.e, to each wavelet subband) $\mathbf{w} = [w_1, ..., w_n]$
   - a vector of significance measures called *significance map*: $\mathbf{m} = [m_1, ..., m_n]$ and
   - a vector of binary labels (hidden variables) $\mathbf{x} = [x_1, ..., x_L]$ called *mask*.
2. Impose a MRF prior on masks and shrink each wavelet coefficient according to probability that it presents a significant signal given the significance map for the whole detail image. In particular,

$$\hat{y}_l = P(X_l = 1|\mathbf{M}=\mathbf{m})w_l. \tag{7}$$

The exact computation of the marginal probability $P(X_l = 1|\mathbf{M} = \mathbf{m})$ is intractable, because it requires the summation of the posterior probabilities $P(\mathbf{X} = \mathbf{x}|\mathbf{M} = \mathbf{m})$ of all possible configurations $\mathbf{x}$ for which $x_l = 1$. In practice one estimates the required probabilities by using a relatively small, but "representative" subset of all possible configurations. Such a representative subset is obtained by *importance sampling*: the probability that a given mask is sampled is proportional to its posterior probability. An estimate of $P(X_l = 1|\mathbf{M} = \mathbf{m})$ is the fraction of all sampled masks for which $x_l = 1$. In [27–30] the Metropolis sampler is used, which starts from a given initial mask and generates from each configuration $\mathbf{x}$, a new, "candidate" mask $\mathbf{x}^c$ by switching the binary label at a random position l. The decision about accepting the change is based on the ratio r of the posterior probabilities of the two configurations $r = P(\mathbf{x}^c|\mathbf{m})/P(\mathbf{x}|\mathbf{m}) = f_{\mathbf{M}|\mathbf{X}}(\mathbf{m}|\mathbf{x}^c)P(\mathbf{x}^c)/(f_{\mathbf{M}|\mathbf{X}}(\mathbf{m}|\mathbf{x})P(\mathbf{x}))$. Under the conditional independence assumption $f_{\mathbf{M}|\mathbf{X}}(\mathbf{m}|\mathbf{x}) = \prod_l f(m_l|x_l)$ the posterior probability ratio reduces to

$$r = \frac{f_{M_l|X_l}(m_l|x_l^c)}{f_{M_l|X_l}(m_l|x_l)}\exp\Big(\sum_{C\in\mathcal{C}_l} V_C(\mathbf{x}) - \sum_{C\in\mathcal{C}_l} V_C(\mathbf{x}^c)\Big). \tag{8}$$

where $\mathcal{C}_l$ is the set of cliques that contain pixel l. When $r > 1$ the candidate $\mathbf{x}^c$ is accepted and if $r < 1$, the change is accepted with probability r. In practice, ten iterations suffice to estimate $P(X_l = 1|\mathbf{M} = \mathbf{m})$ from (7).

3.2.1. Significance measures and their statistics

A significance measure m_l is supposed to tell us how significant the wavelet coefficient w_l is, i.e., to give an indication how likely is it that w_l represents an actual discontinuity rather than being dominated by noise. An obvious and simple choice is the coefficient magnitude $m_l = \omega_l = |w_l|$. If "significant" noise-free coefficient is defined as $|y_l| > T$, where T is some threshold, then $f_{\Omega_l|X_l}(\omega_l|0)$ and $f_{\Omega_l|X_l}(\omega_l|1)$ follow from the conditional densities in Fig. 2(b) as follows: $f_{W_l|X_l}(w_l|0) = f_{Y_l|X_l}(y_l|0) * N(0,\sigma^2)$ and $f_{W_l|X_l}(w_l|1) = f_{Y_l|X_l}(y_l|1) * N(0,\sigma^2)$; $f_{\Omega_l|X_l}(\omega_l|x_l) = 2f_{W_l|X_l}(w_l|x_l)$, $\omega_l > 0$. An illustration of these densities, for different σ, is in Fig. 5.

One can also define a significance measure based on the propagation of the wavelet coefficients across scales: it is well known that the coefficients that die out swiftly as the scale increases are likely to represent noise [49]. In this respect, m_l can be defined as an estimate of the local Lipschitz regularity [27, 48] or as an interscale product [50] at the corresponding spatial position. A rough estimate of the local Lipschitz regularity at position l is *Average Cone Ratio* (ACR) [30]. If we denote by $C(j,l)$ the discrete set of wavelet coefficients at the resolution scale 2^j, which belong to the directional *cone of influence* (Fig. 6(a)) of the spatial position l, then ACR between the scales 2^n and 2^k, $0 < n < k$ is

$$\beta_{n\to k,\ l} \triangleq \log_2\Big(\frac{1}{k-n}\sum_{j=n}^{k-1}\frac{|I_{j+1,l}|}{|I_{j,l}|}\Big), \quad I_{j,l} \triangleq \sum_{m\in C(j,l)} |w_{j,m}|, \tag{9}$$

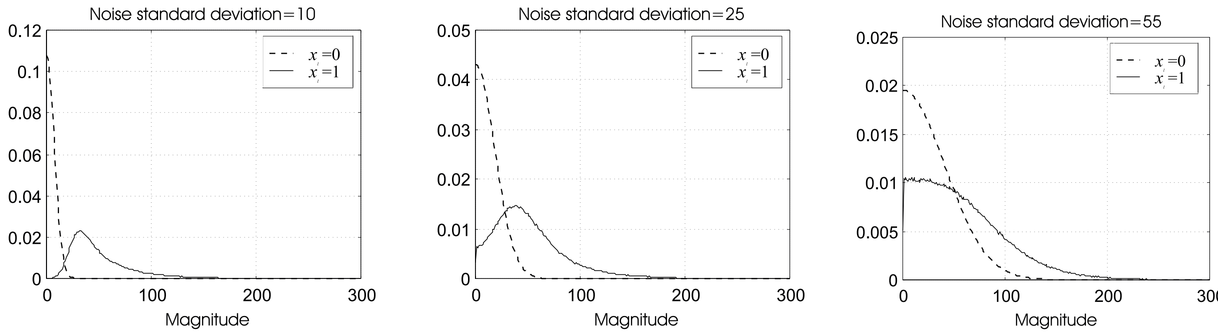


Figure 5. Empirical conditional densities of coefficient magnitudes from [30] that can be also derived analytically from the prior in Fig. 2(b).

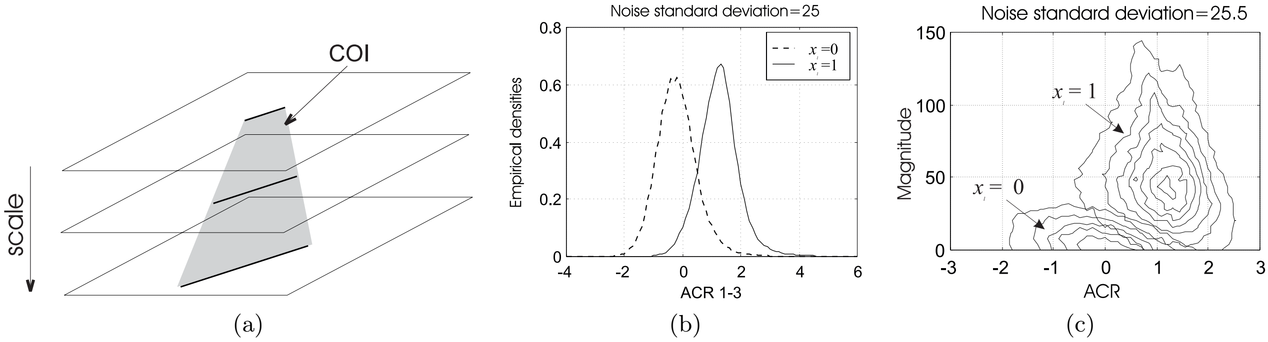


Figure 6. (a) Directional cone of influence in [30,48]. It is support of wavelets in different scales with direction indicated by the wavelet transform angle at a given point. (b) Conditional densities of ACR $\beta_{1\to3,\ l}$ from [30]. (c) Contour plots of joint conditional densities $f_{M_l|X_l}(m_l|0)$ and $f_{M_l|X_l}(m_l|1)$, for $m_l = (|w_l|, \beta_{1\to3,\ l})$ from [30].

and it presents an estimate of $\alpha + 1$, where α is the local Lipschitz regularity (for details, see [30,48]). Fig. 6(b) illustrates the conditional densities of ACR, given noise and given useful signal. In [30], a joint significance measure $m_{j,l} = (|w_l|, \beta_{1\to j+1,\ l})$ was defined. Its empirical conditional densities $f_{M_l|X_l}(m_l|0)$ and $f_{M_l|X_l}(m_l|1)$, illustrated in Fig. 6(c), were shown to be approximated well by the product of the corresponding one dimensional densities from Fig. 5 and Fig. 6(b).

3.2.2. Specification of the MRF prior

A number of different MRF models [47] can be used to express the prior mask probability $P(\mathbf{X} = \mathbf{x})$, but the complexity of realization is an important thing to bear in mind. In [27], an isotropic MRF model, with the second order neighborhood was used. An *anisotropic* MRF model of [30] is slightly more complex but it preserves image details significantly better. The idea behind this model is the following: for each spatial position l, define a given number of oriented *sub-neighborhoods*, which contain possible micro-edges centered at the position l. The label value $x_l = 1$ (*edge label*) should be assigned a high probability if *any* of the oriented sub-neighborhoods indicates the existence of an edge element in a certain direction. On the contrary, the *non-edge* label should be assigned a high probability only if *no one* of the sub-neighborhoods indicates the existence of a such edge element. The sub-neighborhoods $N_{l,i}$, $1 \le i \le 5$ are shown in Fig. 7: each $N_{l,i}$ contains four neighbors of the central pixel l. The expression $\sum_{C\in\mathcal{C}_l} V_C(\mathbf{x})$ that appears in (3.2), for the model of [30] with the label set $x_l \in \{-1, 1\}$ becomes

$$\sum_{C\in\mathcal{C}_l} V_C(\mathbf{x}) = -\gamma\ x_l\ \max_i \Big\{ \sum_{k\in N_{l,i}} x_k \Big\}, \tag{10}$$

where γ is a positive constant. Fig. 8 illustrates the operation of the Metropolis sampler with this prior and Fig. 9 compares image denoising using the above described MRF-based approach to Wiener filtering.

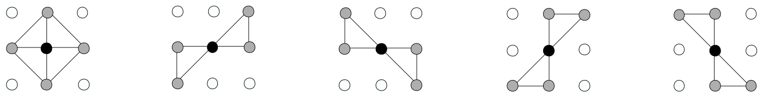


Figure 7. The sub-neighborhoods in the MRF model of [30].

Figure 8. Left to right: noisy image, initial mask and the results of the first three iterations of the Metropolis sampler using the joint conditional model from Fig. 6(c) and the MRF prior with clique potentials in Eq (10).

3.2.3. An alternative to stochastic sampling

As an alternative to the shrinkage rule (7) consider now

$$\hat{y}_l = P(X_l = 1|\mathbf{M} = \mathbf{m}, \mathbf{X}_{L\backslash l} = \hat{\mathbf{x}}_{L\backslash l})w_l. \tag{11}$$

It was shown in [51] that contrasting $P(X_l = 1|\mathbf{M} = \mathbf{m})$, the marginal probability $P(X_l = 1|\mathbf{M} = \mathbf{m}, \mathbf{X}_{Ll} = \hat{\mathbf{x}}_{Ll})$ which is conditioned not only on the significance map $\mathbf{m}$ but also on the estimated labels at all positions except l, can be expressed as a closed form function of m_l and the neighboring labels $\hat{\mathbf{x}}_{\partial l}$. In particular, if the conditional independence $f_{\mathbf{M}|\mathbf{X}}(\mathbf{m}|\mathbf{x}) = \prod_l f_{M_l|X_l}(m_l|x_l)$ holds (as was assumed in Section 3.2 as well) then [47]

$$P(X_l = x_l|\mathbf{M} = \mathbf{m}, \mathbf{X}_{L\backslash l} = \mathbf{x}_{L\backslash l}) = Ap_{M_l|X_l}(m_l|x_l)P(X_l = x_l|\mathbf{X}_{\partial l} = \mathbf{x}_{\partial l}), \tag{12}$$

where A does not depend on x_l. Using (12) and observing that $P(X_l = 0|\mathbf{m}, \hat{\mathbf{x}}_{L\backslash l}) + P(X_l = 1|\mathbf{m}, \hat{\mathbf{x}}_{L\backslash l}) = 1$, the shrinkage rule (11) becomes [51]

$$\hat{y}_l = \frac{\xi_l \mu_l}{1 + \xi_l \mu_l} w_l, \tag{13}$$

where ξ_l is the likelihood ratio and μ_l is the ratio of prior probabilities:

$$\xi_l = \frac{f_{M_l|X_k}(m_l|1)}{f_{M_l|X_l}(m_l|0)} \quad \text{and} \quad \mu_l = \frac{P(X_l = 1|\hat{\mathbf{x}}_{\partial l})}{P(X_l = 0|\hat{\mathbf{x}}_{\partial l})} = \frac{P(X_l = 1|t_l)}{P(X_l = 0|t_l)}, \tag{14}$$

where t_l is a function of label estimates from $\hat{\mathbf{x}}_{\partial l} = \{\hat{x}_k : k \in \partial l\}$. For the isotropic auto-logistic MRF model with pairwise cliques only, one can show that $\mu_l = \exp(\gamma t_l)$, with $t_l = \sum_{k\in\partial l}(2\hat{x}_k - 1)$.

In this approach, instead of using the stochastic sampling, the mask $\hat{\mathbf{x}}$ can be estimated using a fast suboptimum method like *iterated conditional modes* [47] in a few iterations only. The measurement t_l can be interpreted as a *local spatial activity indicator*, which changes the amount of smoothing for a given significance measure depending on the presence of edge-components in a given neighborhood.

4. LOW COMPLEXITY METHODS, HEURISTICS AND EMPIRICAL ESTIMATION

Many authors have used a local measurement such as the locally averaged coefficient magnitude or the local variance in order to refine thresholding [31] and Wiener based estimators (see [32] and the references therein). Here we treat a class of related methods that use the estimator of the form (13), but now beyond MRFs and with more general types of local spatial activity indicators. These algorithms still fit in a Bayesian framework,

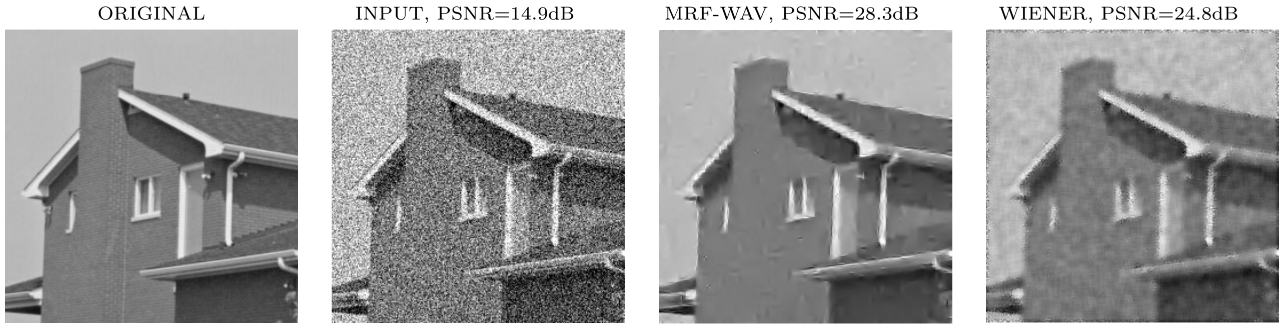


Figure 9. A result of a wavelet domain MRF method [30] (MRF-WAV) compared to spatially adaptive Wiener filter.

but they also involve *heuristics* and *empirical estimation* of the data distributions. Consider a variant of the shrinkage rule (13) with

$$\xi_l = \frac{f_{\Omega_l|X_k}(\omega_l|1)}{f_{\Omega_l|X_l}(\omega_l|0)} \quad \text{and} \quad \mu_l = \frac{P(X_l = 1|z_l)}{P(X_l = 0|z_l)}. \tag{15}$$

where m_l from (14) is now chosen as the coefficient magnitude, denoted by ω_l and where a discrete local spatial activity indicator (LSAI) t_l from (14) is now generalized by z_l. Here z_l denotes an arbitrary, but well chosen function of the neighboring labels (*discrete* LSAI) or a function of the neighboring coefficients $\{w_k : k \in \partial l\}$ (*continuous* LSAI). With this formulation, (13) and (15) provide a heuristically appealing and flexible framework for constructing different denoising methods that are adapted to the data statistics and to the local spatial context, and that have proved effective in different applications including medical imaging [37] and remote sensing [52]. A theoretical motivation also exist: related estimators are widely used in spectral amplitude estimation of speech and image signals [53] where they were also motivated in terms of optimum simultaneous detection and estimation of signals from noise [54].

4.1. Empirical density estimation

In some cases one can develop the estimator given by (13) and (15) analytically, starting e.g., from the mixture prior in Fig. 2(b) (an example is in [39]). In other cases where, e.g., different noise types are considered, or when the notion of "significant" image features is subject to expert-interaction (which can be advantageous in medical images) or in cases where conditional densities of some arbitrary defined local spatial activity indicators are required, the empirical estimation is required. In [37,52] an empirical density estimation is based on a preliminary coefficient classification. In particular, a non-decimated transform is used and the positions of "significant" coefficients are estimated using a coarse-to fine procedure: already processed, coarser detail coefficients $\hat{y}_{l,j+1}$ at the scale 2^{j+1}, are used to better detect the important ones at the scale 2^j. For each orientation we have

$$\hat{x}_{j,l} = \begin{cases} 0, & \text{when } |w_{l,j}||\hat{y}_{l,j+1}| < (K\hat{\sigma}_j)^2, \\ 1, & \text{when } |w_{l,j}||\hat{y}_{l,j+1}| \geq (K\hat{\sigma}_j)^2, \end{cases} \tag{16}$$

where $\hat{\sigma}_j$ is an estimate of the noise standard deviation at the resolution scale 2^j and K is a parameter, which controls the notion of the "signal of interest". K can be set to a fixed value or in some sensitive applications, like medical ultrasound, a user interaction may be preferred. Having the estimate $\hat{\mathbf{x}} = \{\hat{x}_1 ... \hat{x}_n\}$, let

$$S_0 = \{l : \hat{x}_l = 0\} \quad \text{and} \quad S_1 = \{l : \hat{x}_l = 1\}. \tag{17}$$

The normalized histograms of $\{\omega_l : l \in S_0\}$ and $\{\omega_l : l \in S_1\}$ are the empirical estimates of the densities $f_{\Omega_l|X_l}(\omega_l|0)$ and $f_{\Omega_l|X_l}(\omega_l|1)$, respectively. From thereon, at least two strategies are possible, as it is depicted in Fig. 10. If the functional form of the involved densities is unknown, one can perform a *piece-wise linear fitting of the log-ratio* (Section 4.2). Otherwise, one can apply the *maximum likelihood estimation* of the model parameters from the corresponding histograms (Section 4.3).

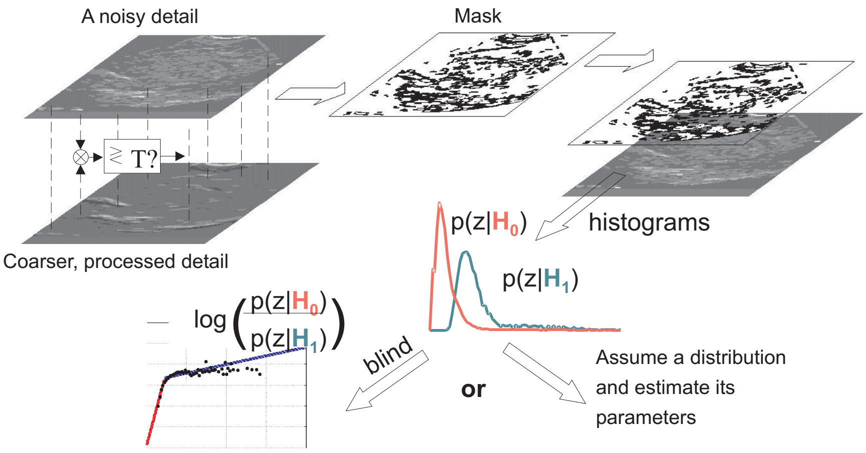


Figure 10. Empirical density estimation using a preliminary coefficient classification.

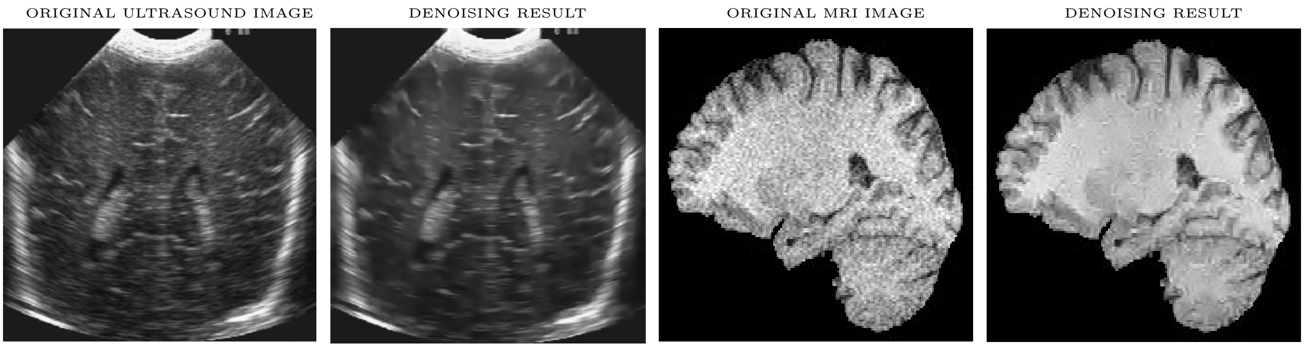


Figure 11. Denoising ultrasound and magnetic resonance images (MRI) using the method of [37].

4.2. An algorithm for medical images

A representative of this approach is the algorithm of [37]. The local spatial activity indicator is there defined as the locally averaged coefficient magnitude and μ_l from (15) is expressed as

$$\mu_l = \frac{P(X_l = 1|z_l)}{P(X_l = 0|z_l)} = \frac{f_{Z_l|X_l}(z_l|1)P(X_l = 1)}{f_{Z_l|X_l}(z_l|0)P(X_l = 0)} \tag{18}$$

yielding from (13)

$$\hat{y}_l = \frac{\rho\xi_l\eta_l}{1 + \rho\xi_l\eta_l} w_l, \tag{19}$$

where

$$\xi_l = \frac{f_{\Omega_l|X_k}(\omega_l|1)}{f_{\Omega_l|X_l}(\omega_l|0)}, \eta_l = \frac{f_{Z_l|X_k}(z_l|1)}{f_{Z_l|X_l}(z_l|0)}, \quad \text{and} \quad \rho = \frac{P(X_l = 1)}{P(X_l = 0)}. \tag{20}$$

The likelihood ratios ξ_l and η_l are estimated from the noisy image using (16) followed by a piece-wise linear fitting of the log-likelihood ratios (see Fig. 10). $P(X_l = 1)$ is estimated by a fraction of estimated labels 1, yielding $\rho = \sum_{l=1}^{n} \hat{x}_l / (n - \sum_{l=1}^{n} \hat{x}_l)$, where n is the number of the coefficients in a given subband. Fig. 11 illustrates the application of this method to medical ultrasound brain images and to magnetic resonance images of human brain.

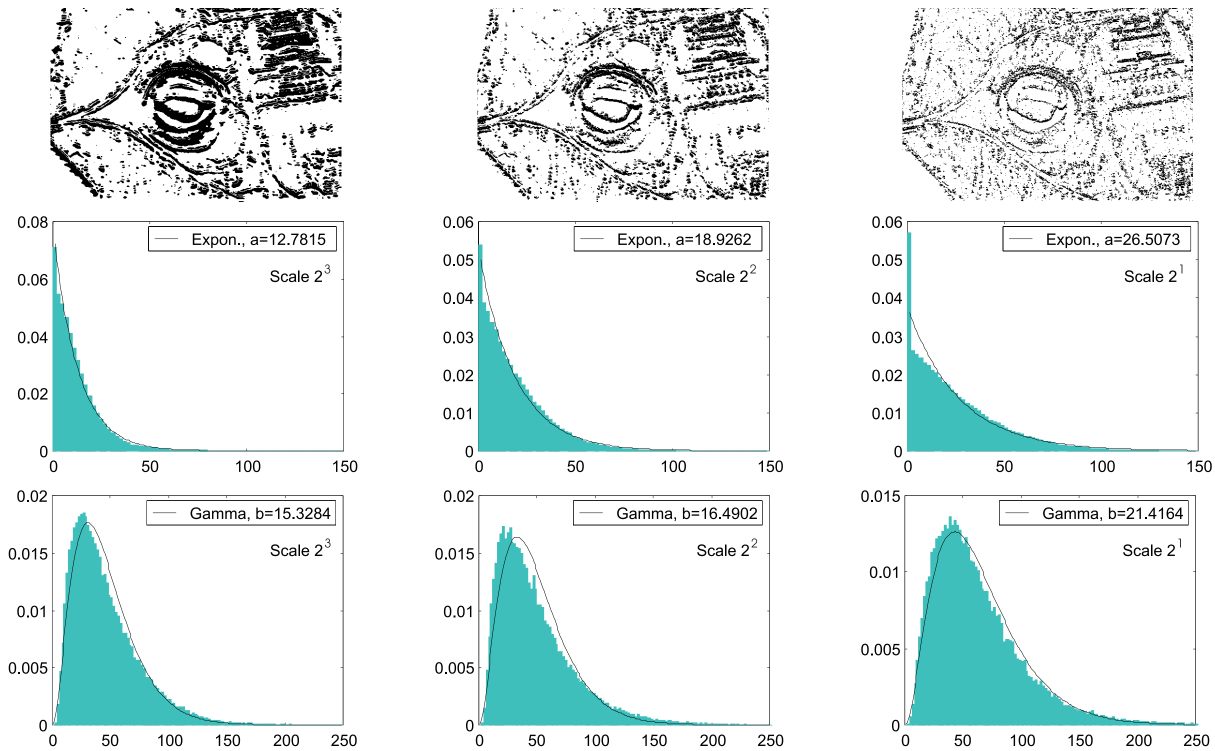


Figure 12. Empirical density estimation of [52] for the SAR image from Fig. 13. Top: detected masks. Middle row: empirical histograms and fitted models for $f_{\Omega|X}(\omega|0)$. Bottom: empirical histograms and fitted models for $f_{\Omega|X}(\omega|1)$.

4.3. Algorithm for SAR image despeckling

A related algorithm of [52] for despeckling Synthetic Aperture Radar (SAR) images employs the estimator (13) using the discrete local spatial activity from (14). As the previous one, this method employs an empirical density estimation according to (16) and Fig. 10, but using now *functional forms* of the involved densities. It was observed in [52] that in SAR images, the coefficient magnitudes dominated by speckle noise follow well a scaled exponential density, while those dominated by image transitions follow well scaled Gamma densities, i.e.,

$$f_{\Omega_l|X_l}(\omega|0) \simeq (1/a)\exp(-\omega/a) \quad \text{and} \qquad (21)$$
$$f_{\Omega_l|X_l}(\omega|1) \simeq (1/2b)(\omega/b)^2\exp(-\omega/b).$$

It can be shown (see [55]) that the maximum likelihood estimates of these parameters are: $\hat{a} = (1/N_0)\sum_{i\in S_0}\omega_i$ and $\hat{b} = (1/3N_1)\sum_{i\in S_1}\omega_i$, where the sets S_0 and S_1 are defined in (17) and N_0 and N_1 denote the cardinalities of S_0 and S_1, respectively. An example in Fig. 12 illustrates the coarse-to-fine preliminary coefficient classification according to (16) and the the estimated conditional densities of the coefficient magnitudes. Denoising results in Fig. 13 and Fig. 14 demonstrate that this method preserves the point scatterers remarkably well and that it visually outperforms the Gamma map filter [56], which is one of the best state of the art speckle filters.

5. CONCLUSIONS

We discussed different multiscale statistical image models in the framework of Bayesian image denoising. The choice of a marginal prior is a crucial step for designing a Bayesian denoising method. We discussed different heavy tailed and mixture priors from the viewpoint of complexity and flexibility in applications. We remarked that the use of mixtures of two distributions (2) seems particularly attractive for applications where in addition to the noisy data some other sources can be used to locate the image edges (like in case of multivalued images).

Figure 13. Original SAR image (left) and the result of the wavelet domain filter [52] (right)

Figure 14. Original SAR image (left), wavelet based filter [52] (middle) and the Gamma MAP filter [56] (right).

HMT and MRF approaches were briefly outlined, where we through some new and original illustrations emphasized the main conceptual differences and similarities between these approaches and where we also discussed how these models build on the chosen marginal priors. Finally, we reviewed a class of low-complexity locally adaptive methods within a new, unifying framework, drawing also some parallels between these methods and the MRF-based ones. Potentials of denoising methods that employ local spatial activity indicators and the empirical density estimation was illustrated on different real world images: ultrasound, MRI and SAR.

REFERENCES

1. A. Willsky, "Multiresolution markov models for signal and image processing," *Proceedings of the IEEE* **90**, pp. 1396–1458, Aug. 2002.
2. I. Daubechies, S. Mallat, and A. Willsky, "Special issue on wavelet transforms and multiresolution signal analysis," *IEEE Trans. Information Theory* **38**, Mar. 1992.
3. H. Krim, W. Willinger, A. Juditski, and D. Tse, "Special issue on multiscale statistical signal analysis and its applications," *IEEE Trans. Information Theory* **45**, Apr. 1999.
4. F.-L. Starck and A. Bijaoui, *Multiscale image processing and data analysis*, Cambridge Univer. Press, Cambridge, UK, 1998.
5. I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.

6. S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. and Machine Intel.* **11**(7), pp. 674–693, 1989.
7. S. Mallat, *A wavelet tour of signal processing*, Academic Press, London, 1998.
8. M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, 1995.
9. D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory* **41**, pp. 613–627, May 1995.
10. D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.* **90**, pp. 1200–1224, Dec. 1995.
11. B. Vidakovic, "Nonlinear wavelet shrinkage with bayes rules and bayes factors," *J. of the American Statistical Association* **93**, pp. 173–179, 1998.
12. B. Vidakovic, "Wavelet-based nonparametric bayes methods," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, D. D. Dey, P. Müller, and D. Sinha, eds., *Lecture Notes in Statistics* **133**, pp. 133–155, Springer Verlag, New York, 1998.
13. D. Leporini, J. C. Pasquet, and H. Krim, "Best basis representation with prior statistical models," in *Lecture Notes in Statistics*, P. M ed.
14. S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Proc.* **9**, pp. 1532–1546, Sept. 2000.
15. F. Abramovich, T. Sapatinas, and B. Silverman, "Wavelet thresholding via a bayesian approach," *J. of the Royal Statist. Society B* **60**, pp. 725–749, 1998.
16. H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive bayesian wavelet shrinkage," *J. of the Amer. Statist. Assoc* **92**, pp. 1413–1421, 1997.
17. M. Clyde, G. Parmigiani, and B. Vidakovic, "Multiple shrinkage and subset selection in wavelets," *Biometrika* **85**(2), pp. 391–401, 1998.
18. E. P. Simoncelli and E. H. Adelson, "Noise removal via bayesian wavelet coring," in *Proc. IEEE Internat. Conf. Image Proc. ICIP*, (Lausanne, Switzerland, 1996).
19. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors," *IEEE Trans. Inform. Theory* **45**, pp. 909–919, Apr. 1999.
20. L. Şendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet -based denoising exploiting interscale dependency," *IEEE Trans. Signal Proc.* **50**, pp. 2744 –2756, Nov. 2002.
21. E. P. Simoncelli, "Modeling the joint statistics of image in the wavelet domain," in *Proc. SPIE Conf. on Wavelet Applications in Signal and Image Processing VII*, (Denver, CO).
22. M. S. Crouse, R. D. Nowak, and R. G. Baranuik, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. Signal Proc.* **46**(4), pp. 886–902, 1998.
23. J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree structured image modeling using wavelet-domain hidden markov models," *IEEE Trans. Image Proc.* **10**(7), pp. 1056–1068, 2001.
24. G. Fan and X. G. Xia, "Image denoising using local contextual hidden markov model in the wavelet domain," *IEEE Signal Processing Letters* **8**, pp. 125–128, May 2001.
25. M. Wainwright, E. Simoncelli, and A. Willsky, "Random cascades on wavelet trees and their use in modeling natural images," *Appl. Comput. Harmon. Anal.* **11**, pp. 89–123, 2001.
26. R. Nowak, "Multiscale hidden markov models for bayesian image analysis," in *Bayesian inference in wavelet based models*, P. Müller and B. Vidakovic, eds., New York: Springer Verlag, 1999.
27. M. Malfait and D. Roose, "Wavelet-based image denoising using a markov random field a priori model," *IEEE Trans. Image processing* **6**(4), pp. 549–565, 1997.
28. M. Jansen and A. Bultheel, "Geometrical priors for noisefree wavelet coefficients in image denoising," in *Bayesian inference in wavelet based models*, P. Müller and B. B. Vidakovic, eds., *Lecture Notes in Statistics* **141**, pp. 223–242, Springer Verlag, 1999.
29. M. Jansen and A. Bultheel, "Empirical bayes approach to improve wavelet thresholding for image noise reduction," *J. Amer. Stat. Assoc.* **96**(454), pp. 629–639, 2001.
30. A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for wavelet based bayesian image denoising," *IEEE Trans. Image Proc* **11**, pp. 545–557, May 2002.
31. S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Proc.* **9**, pp. 1522–1531, Sept. 2000.

32. M. K. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Proc. Lett.* **6**, pp. 300–303, Dec. 1999.
33. V. Strela, J. Portilla, and E. P. Simoncelli, "Image denoising using a local gaussian scale mixture model in the wavelet domain," in *Proc. SPIE, 45th Annual Meeting*, (San Diego, July 2000.).
34. J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Adaptive wiener denoising using a gaussian scale mixture model in the wavelet domain," in *Proc. IEEE Internat. Conf. on Image Proc.*, (Thessaloniki, Greece, Oct. 2001.).
35. J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using gaussian scale mixtures in the wavelet domain," *IEEE Trans. Image Proc. (accepted)* .
36. X. Li and M. Orchard, "Spatially adaptive denoising under overcomplete expansion," in *Proc. IEEE Internat. Conf. on Image Proc.*, (Vancouver, Canada), Sept. 2000.
37. A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Trans. Medical Imaging* **22**(3), pp. 323–331, 2003.
38. M. Hansen and B. Yu, "Wavelet thresholding via mdl for natural images," *IEEE Trans. Inform. Theory* **46**, pp. 1778–1788, Aug. 2000.
39. A. Pižurica and W. Philips, "Adaptive probabilistic wavelet shrinkage for image denoising." Technical Report, June 2003, (submitted to *IEEE Trans. Image Proc.*).
40. C. Bouman and K. Sauer, "A generalized gaussian image model for edge preserving map estimation," *IEEE Trans. Signal Proc.* **2**, pp. 296–310, 1993.
41. S. Foucher, G. Bénié, and J. Boucher, "Multiscale map filtering of sar images," *IEEE Trans. Image Proc.* **10**, pp. 49–60, Jan. 2001.
42. A. Achim, A. Bezerianos, and P. Tsakalides, "Novel bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Medical Imaging* **20**, pp. 772–783, Aug. 2001.
43. J. Liu and P. Moulin, "Analysis of interscale and intrascale dependencies between image wavelet coefficients," in *Proc. Int. Conf. on Image Proc., ICIP*, (Vancouver, Canada, Sep. 2000.).
44. M. Basseville, A. Benveniste, K. Chou, S. Golden, R. Nikoukhah, and A. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory* **38**, pp. 766–784, Mar. 1992.
45. M. Banham and A. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration,"
46. M. Luettgen, W. Karl, A. Willsky, and R. Tenney, "Multiscale representations of markov random fields," *IEEE Trans. Signal Proc.* **41**(12), pp. 3377–3396, 1993.
47. S. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, 1995.
48. T.-C. Hsung, D.-K. Lun, and W.-C. Siu, "Denoising by singularity detection," *IEEE Trans. Signal Proc.* **47**, pp. 3139–3144, Nov. 1999.
49. S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Information Theory* **38**, pp. 617–643, Mar. 1992.
50. Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu, "Wavelet transform domain filters: a spatially selective noise filtration technique," *IEEE Trans. Image Proc.* **3**, pp. 747–758, Nov. 1994.
51. A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A wavelet-based image denoising technique using spatial priors,"
52. A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "Despeckling sar images using wavelets and a new class of adaptive shrinkage estimators,"
53. T. Aach and D. Kunz, "Anisotropic spectral magnitude estimation filters for noise reduction and image enhancement," in *Proc. IEEE Internat. Conf. on Image Proc. ICIP96*, pp. 335–338, (Lausanne, Switzerland, 16-19 Sep, 1996.).
54. D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inform. Theory* **14**, pp. 434–443, May 1968.
55. A. Pižurica, *Image Denoising Using Wavelets and Spatial Context Modeling*. PhD thesis, Ghent University, Belgium, 2002.
56. A. Lopes, R. Touzi, and E. Nezry, "Structure detection and statistical adaptive speckle filtering in sar images," *Int. J. Remote Sensing* **14**(9), pp. 1735–1758, 1990.