THE APPLICATION OF MARKOV RANDOM FIELD MODELS TO WAVELET-BASED IMAGE DENOISING

A. PIZURICA, W. PHILIPS, I. LEMAHIEU AND M. ACHEROY

Abstract. In this chapter, we address the use of Markov Random Field (MRF) prior models in wavelet based image denoising. Two different approaches are considered: the maximum a posteriori (MAP) estimation of the wavelet coefficients using MRF priors, and the MAP estimation of significant edges in detail images. The second approach can be incorporated into many different wavelet based denoising schemes; in particular, we discuss its use in wavelet shrinkage techniques. Practical results demonstrate that this approach is very efficient for suppressing different kinds of natural noise, which makes it attractive for different applications.

1. Introduction. The wavelet representation has become very attractive for image denoising because it naturally facilitates the construction of *spatially adaptive* algorithms, which adapt to real discontinuities in the image and smooth noise only in the areas where no significant details are present. The wavelet decomposition enables one to analyze different image details at different resolution scales. At each resolution scale the wavelet coefficients can be organized into *detail images*, which represent a bandpass content of the image in the specific (horizontal, vertical or diagonal) spatial direction. Large wavelet coefficients occur exclusively in the areas of image discontinuities, such as edges, peaks, corners, etc. Due to this property, noise can be effectively suppressed even by simple thresholding [25] of the wavelet coefficients.

Advanced wavelet based denoising techniques make use of interscale dependencies [18, 36, 48, 66], intrascale (spatial) dependencies [42, 56, 62] or both inter and intrascale dependencies [28, 29, 51, 58] between wavelet coefficients. Spatial dependencies in detail images, as well as those among pixel values in general, can be quantitatively expressed using the appropriate image models.

Markov Random Field (MRF) models are widely used in image processing, especially for texture modeling and classification [1, 14, 19], image restoration [32, 33], and segmentation of noisy and textured images [8, 17, 22, 43]. The classical concepts of single-resolution image restoration using MRFs and the maximum a posteriori (MAP) estimation were established in [32]. The extension of this approach to a multigrid framework [33], has proved its advantages: better results are obtained, and at the same time, the computation time is reduced. The implementation of these ideas in the wavelet domain is a research field with great expectations, but also with a number of different possibilities and open questions.

For the wavelet representation, multiscale or hierarchical MRF models, like the one in [40], seem to be the most natural choice. These models enable a complete modeling of the wavelet coefficients of images, including both interscale and spatial dependencies. On the other hand, such models call for a big memory and a long computation time. Their application in the wavelet domain is not well studied yet. Instead, many wavelet based techniques apply hidden Markov tree (HMT) models [18, 59]. For images, these models are defined on a quad-tree structure, and lead to much faster algorithms than multiscale MRF. However, HMT models do not capture the spatial dependencies between wavelet coefficients, but only the evolution of their magnitudes through resolution scales. Such models are naturally related with the wavelet representation of one-dimensional signals, but do not encounter the specific properties of images (the spatial context). To overcome this disadvantage, so-called contextual HMT model was recently proposed [28, 29]. It combines a simple local

spatial model into a classical HMT. For encoding the prior knowledge about spatial dependencies, MRF models remain more powerful. Our interest in this chapter is the application of the MRF models in the wavelet domain.

An attractive approach in this respect was proposed in [51]. It uses a binary MRF to encode the prior knowledge about the "geometry" of detail images, and applies the MAP estimation of the positions of useful edges. The interscale dependencies are taken into account through a specific data distribution model. Related approaches were later proposed, e.g., in [57, 58]. These techniques apply a Bayesian shrinkage of the wavelet coefficients, using MRF priors for their spatial clustering. Practically, the shrinkage factor applied to a given wavelet coefficient depends not only on the measurement at the corresponding spatial position, but it also depends on the spatial context. One should note that this approach differs significantly from [35]. There, the authors also note that significant wavelet coefficients are clustered in space, but apply it to a different problem: to better estimate an "optimum" threshold value, which is constant for a whole detail image. Using such a fixed threshold one cannot preserve edge clusters of wavelet coefficients, and suppress at the same time larger, but isolated ones. And precisely this is enabled by using MRF priors.

In this chapter, we have three main objectives. The first one is to present the basic concepts of the wavelet based image denoising using MRF prior models in a comprehensive way, suitable for non-experts in this field. The second objective is to emphasize its various applications. Indeed, the Bayesian shrinkage using MRF priors proves to be very useful for different kinds of natural images. In particular, we demonstrate its application to infrared images (humanitarian demining), ultrasound images (medicine) and radar images (remote sensing). Finally, our third objective is to discuss a general framework of using the MAP-MRF denoising in the wavelet domain. The use of binary MRF prior models is only one possibility in this respect. Instead, one can implement the classical MAP-MRF estimation in the fashion of [32] to the wavelet coefficients. We do not propose any practical algorithm of this kind, but we do discuss some possibilities and point some advantages and specific problems.

This chapter is organized as follows. In Section 2, we explain briefly the wavelet transform and the related multiresolution representation of signals. The basic concepts of wavelet based image denoising are introduced in Section 3. We discuss in more detail some approaches that can be further extended or combined with MRF priors. Next, in Section 4, we address the use of spatial context in image denoising. The relationship between Markov and Gibbs random fields is discussed and some often used MRF models are presented. The main subject is treated in Section 5, where we apply MRF models to wavelet based image denoising. We use the MAP approach and apply it in two different directions: (i) to estimate the "actual" values of the wavelet coefficients and (ii) to estimate masks, which indicate the positions of meaningful wavelet coefficients. The computational aspects for both of these cases are discussed. In Section 6, we present some results and discuss different applications.

2. The Wavelet Representation. In signal processing theory and applications the representation of signals plays a fundamental role. We need to represent signals in a way that facilitates the extraction of certain characteristics or features that are important for a specific problem (e.g., the occurrence of abrupt changes in time or edges and different textures in images). It is essential for the chosen representation to be computed quickly, because otherwise it has little practical importance. A common way to obtain a specific representation is to decompose a signal $f(x)$ into elementary "atoms" or building blocks $\psi_i(x)$ that are simple waveforms $f(x) = \sum_i \alpha_i \psi_i(x)$.

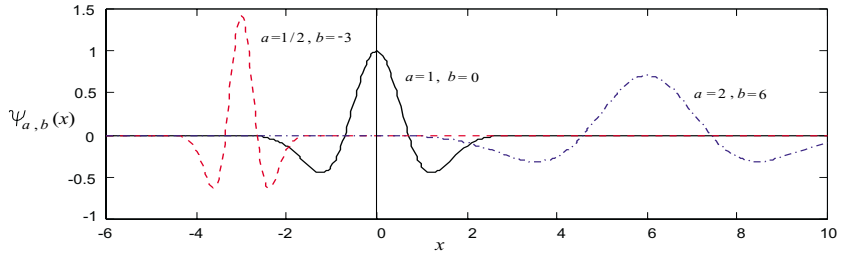


FIGURE 2.1. *A few wavelets obtained from the Mexican hat mother wavelet* $\psi(x) = (1-2x^2)e^{-x^2}$.

For many applications (like compression and denoising) it is extremely important to have a representation with good *approximation properties*, i.e., with the ability to represent a signal at a given level of accuracy with as few building blocks as possible. Another important point is that details in signals or images that we want to analyze may have many different sizes. It is generally difficult to find one single and optimal resolution scale to analyze the complete signal. The power of a *multiresolution* representation is that it enables one to analyze different details at different resolution scales. All these advantages are offered by the wavelet decomposition, which has therefore become a widely used representation in signal and image processing.

In one dimensional wavelet representation, the signal is decomposed into a family of wavelets

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi(\frac{x-b}{a}), \qquad a > 0,\ b \in \mathbb{R} \tag{2.1}$$

that are obtained from a unique "mother wavelet" $\psi(x)$, by dilating (scaling) it by a and translating it by b. A mother wavelet $\psi(x)$ is generally a function in the Hilbert space $L_2(\mathbb{R})$ (the space of square integrable functions) that has unit norm $\|\psi(x)\| = 1$, oscillates around the x-axis, averages to zero and has a compact support (it is well localized around $x = 0$). A more formal characterization of wavelet functions is given, e.g., in [16, 20, 50].

In Fig. 2.1, several wavelets are shown that are obtained from the mother wavelet $\psi(x) = (1-2x^2)e^{-x^2}$; this wavelet is the second derivative of a Gaussian function and is called the *Mexican hat*. At small resolution scales, the localization of wavelets in signal domain gets finer, which is appropriate for analyzing fine, high frequency details. At coarser scales the wavelets are more elongated in the signal domain and better localized in the frequency domain, which is appropriate for analyzing the slow changes in a signal. The wavelet representation can be implemented using the *continuous* wavelet transform, using wavelet *frames* or using wavelet *bases*. Each of these representations has its advantages for specific applications. We shall use for denoising the frame representation, but first we start with a general brief overview.

2.1. Continuous wavelet transform. The continuous wavelet transform of a signal $f(x)$ is defined as

$$\mathcal{W}f(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(x)\overline{\psi}\Big(\frac{x-b}{a}\Big)dx \ = \ \langle f, \psi_{a,b}\rangle, \tag{2.2}$$

where $\overline{\psi}(x)$ denotes the complex conjugate of $\psi(x)$ and $\langle f, g\rangle$ is the usual notation for the inner product $\int_{-\infty}^{\infty} f(x)\overline{g}(x)dx$. The existence of the inverse transform is

guaranteed if the wavelet is chosen such that $\int_{-\infty}^{\infty} |\hat{\psi}(\omega)|^2/|\omega| d\omega < +\infty$, which implies $\hat{\psi}(0) = 0$. Therefore, $\psi(x)$ can be viewed as an impulse response of a bandpass filter. The continuous wavelet transform is redundant. Its advantages over the discrete wavelet representations are the greater freedom in the choice of a wavelet and the *shift invariance*. An important property of this transform, which is often exploited for denoising, is the following [50]: the rate of increase or decrease of $|\mathcal{W}f(a,x)|$ through resolution scales is directly related to the local *Lipschitz* regularity[1] α of the signal. Formally, if a function $f(x)$ is Lipschitz α at a point x_0, then

$$|\mathcal{W}f(a,x)| \leq A\ a^{\alpha+\frac{1}{2}}, \quad \text{for} \quad |x-x_0| \leq Ka, \tag{2.3}$$

where K is the support of the mother wavelet $\psi(x)$. A region in the scale-space plane (a,x), for which $|x-x_0| \leq Ka$ is called the *cone of influence* (COI) of the point x_0. If a function $f(x)$ has $\alpha > 0$ at a certain point, then the amplitude of the wavelet transform increases through scales within the COI of that point. In the case where $\alpha < 0$ the amplitude of the wavelet transform decreases through scales. This was first in [48] exploited for denoising, assuming that for meaningful signal transitions $\alpha > 0$ and for those that originate from noise $\alpha < 0$.

It is difficult to implement the continuous wavelet transform with a fast algorithm. In practical applications it can be computed only on a discrete grid, which involves sampling in scale and location.

2.2. Wavelet bases. Among the most remarkable achievements in the wavelet theory is the construction of *orthonormal* wavelet bases and the development of fast algorithms [46, 47] for decomposing signals into wavelet basis functions. There exist some wavelets [20] such that the family $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, with $j,k \in \mathbb{Z}$ constitutes an orthonormal basis for the Hilbert space $L^2(\mathbb{R})$, which means that each function f in $L^2(\mathbb{R})$ can be represented as

$$f(x) = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}(x). \tag{2.4}$$

The *wavelet coefficients* in this representation are obtained by sampling the continuous wavelet transform (2.2) at *dyadic* scales $a = \{2^{-j}\}_{j\in\mathbb{Z}}$ and at locations proportional to the scale $b = \{2^{-j}k\}_{j,k\in\mathbb{Z}}$, i.e., $w_{j,k} = \langle f, \psi_{j,k}\rangle = \mathcal{W}f(2^{-j}, 2^{-j}k)$. A greater index j corresponds to a smaller scale 2^{-j} and thus to a finer resolution.

With each mother wavelet $\psi(x)$, a unique *scaling function* $\varphi(x)$ is associated [47], which is often called a "father wavelet". Scaling functions $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$ constitute an orthonormal set for fixed j. In a multiresolution analysis, the decomposition of a signal into the basis $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$ is used to *approximate* a signal at the resolution scale 2^{-j}. The *details* that constitute the difference between approximations at the resolution scale 2^{-j} and the next finer scale $2^{-(j+1)}$ are decomposed into the basis $\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$. This multiresolution concept is depicted in Fig. 2.2. In the engineering terminology, the scaling coefficients $u_{j,k} = \langle f\varphi_{j,k}\rangle$ represent the lowpass content of the signal $f(x)$ and the wavelet coefficients $w_{j,k} = \langle f\psi_{j,k}\rangle$ its bandpass content at a given resolution scale.

Image processing methods make use of two-dimensional wavelets, and commonly three separable wavelets are used: $\psi^{LH}(x,y) \triangleq \varphi(x)\psi(y)$, $\psi^{HL}(x,y) \triangleq \psi(x)\varphi(y)$ and

[1] A function $f(x)$ is said to be Lipschitz $\alpha \geq 0$ at a point x_0 if and only if there exist two positive constants A and h_0, and a polynomial $P_n(x)$ of order $n < \alpha$, such that $|f(x_0+h) - P_n(h)| \leq A|h|^\alpha$, for $h < h_0$. The sup of the set of all α for which this is satisfied is called the Lipschitz regularity.

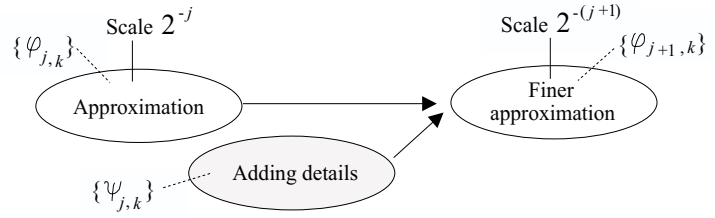


FIGURE 2.2. *The multiresolution concept.*

$\psi^{HH}(x,y) \triangleq \psi(x)\psi(y)$; these wavelets correspond to horizontal, vertical and diagonal details, respectively. In practice a finite number of resolution scales is used and an image $f(x,y)$ is decomposed as the sum

$$f(x,y) = \sum_k u_{n,k}\varphi_{n,k}(x,y) + \\ + \sum_{j=n}^{N-1}\sum_k (w^{LH}_{j,k}\psi^{LH}_{j,k}(x,y) + w^{HL}_{j,k}\psi^{HL}_{j,k}(x,y) + w^{HH}_{j,k}\psi^{HH}_{j,k}(x,y)), \tag{2.5}$$

where $j = n$ corresponds to the coarsest and $j = N$ to the finest resolution scale. To keep the notation short, we have used in (2.5) one index k to represent a *pair* of shifts in x and y directions, i.e., $\psi_{j,k}(x,y) = 2^j\psi(2^jx-k_1, 2^jy-k_2)$, where $k = (k_1,k_2) \in \mathbb{Z}^2$.

The wavelet decomposition (2.5) reorganizes the image content into a set of images: the *lowpass image*, i.e., the set of scaling coefficients $\mathbf{u}_n = \{u_{n,k}\}$, $k \in \mathbb{Z}^2$ that approximate $f(x,y)$ at the coarsest scale, and several *bandpass* or *detail images* $\mathbf{w}^1_j = \{w^{LH}_{j,k}\}, \mathbf{w}^2_j = \{w^{HL}_{j,k}\}$ and $\mathbf{w}^3_j = \{w^{HH}_{j,k}\}$, $k \in \mathbb{Z}^2$ containing details (edges) of different orientations LH, HL, HH and at different resolution scales $n < j < N$.

Wavelet bases provide a nonredundant representation of signals in terms of their wavelet coefficients. An efficient algorithm for *discrete* nonredundant wavelet transform, implemented with filter banks, was proposed in [46] and [47]. Rows and columns of an image are filtered with given lowpass and bandpass filters, and subsequently *downsampled* by two (removing every other sample). Therefore, at each coarser scale the number of samples is reduced by a factor two in each dimension. The discrete nonredundant wavelet representation is widely used for image coding and compression [3, 60], as well as image denoising [4], [23]-[26], [28, 54, 61, 65, 67].

2.3. Wavelet frames. In many practical applications it is better to avoid downsampling at successive resolution scales and to represent a signal with the same number of wavelet coefficients at each scale. Such a redundant representation, using dyadic scales, is obtained by decomposing a signal into a family of wavelets $\psi_{j,k} = 2^{j/2}\psi(2^j(x-k))$. In this representation the wavelet coefficients are samples of the continuous wavelet transform at all integer locations at each dyadic scale, i.e., $w_{j,k} = \mathcal{W}f(2^{-j},k)$. The wavelets $2^{j/2}\psi(2^j(x-k))$ are not linearly independent and thus they do not constitute a basis but a *frame*. A detailed analysis of wavelet frames is presented [16, 20, 50].

In this chapter, we shall use the redundant frame representation, with spline wavelets from [49], in the classical form of (2.5). An example of such a decomposition is given in Fig. 2.3. It is computed using the *à trous* algorithm [49, 50]. This redundant decomposition requires more calculations and calls for bigger memory than the

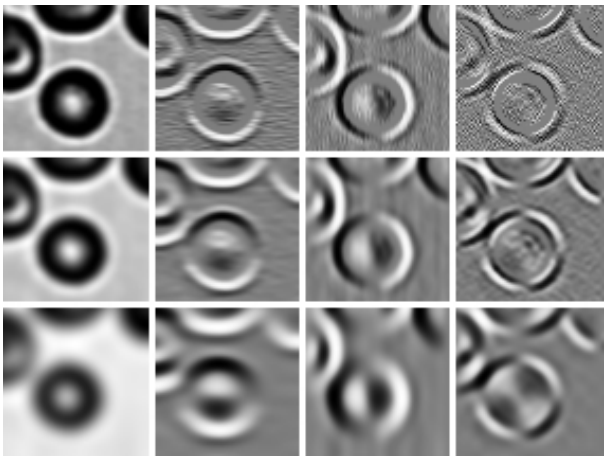


FIGURE 2.3. *An example of the redundant wavelet frame decomposition in three resolution levels. From left to right are represented lowpass images and detail images in LH, HL and HH subbands, respectively.*

decomposition into wavelet bases. However, this is a price worth paying for better solving some practical problems. Several advantages of are listed below.

In a redundant representation, the wavelet coefficients are well correlated across scales. The relation between the wavelet coefficients at the same location in successive resolution levels ($w_{j,k}$ and $w_{j+1,k}$) is a source of additional information that can be exploited to better distinguish useful signal from noise.

The nonredundant representation is not shift invariant. When the signal is shifted its continuous wavelet transform is translated by the same amount, but this is not generally true for the sampled coefficients. To better approach the translation invariance one needs to sample the coefficients at more dense locations. The representation with equally dense sampling at all scales is in this respect better than the wavelet basis representation. The shift invariance is important for pattern recognition purposes, but also for the estimation of local Lipschitz exponents (Section 2.1).

The nonredundant representation is usually implemented for discrete signals or images which size is a power of two, because the number of coefficients is halved in subsequent resolution levels. When this downsampling is avoided, the wavelet transform is equally implemented for arbitrary input sizes.

Finally, for the topic of this chapter the redundant representation is more interesting; spatial dependencies between wavelet coefficients are much higher and more important for denoising than in the case where orthogonal transform is used.

3. Wavelet-Based Image Denoising Methods. The common procedure in wavelet based denoising is the following: (1) Compute the discrete wavelet transform; (2) Modify the wavelet coefficients in order to remove noise and (3) Reconstruct the denoised signal from the modified wavelet coefficients.

We shall use the following notations. Random variables will be denoted by capital letters and their realizations by the corresponding small letters. Boldface capital letters will be used for vectors of random variables and boldface small letters for vectors of realizations. Furthermore, in order to have a more compact representation, we shall omit the indices of the wavelet coefficients that indicate the scale and the orientation, unless in cases where it is explicitly needed. Thus, the wavelet coefficients

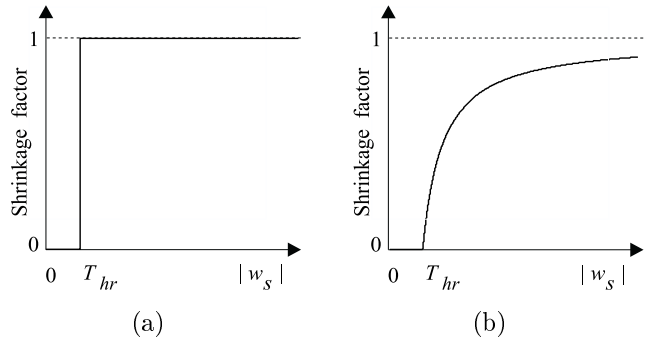


FIGURE 3.1. *Shrinkage factors that multiply the wavelet coefficients in (a) hard-thresholding and (b) soft-thresholding.*

will usually have only one index that corresponds to their spatial location and a given detail image will be represented as $\mathbf{w} = \{w_1, ..., w_n\}$. We assume the following model

$$\mathbf{w} = \mathbf{y} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{y}$ is the unknown "clean" detail image that we want to estimate and $\mathbf{n}$ is the additive white Gaussian noise.

3.1. Thresholding of the wavelet coefficients. In the wavelet domain, the essential information about the signal is contained within relatively few, large coefficients. The large wavelet coefficients occur exclusively in the areas of major "spatial activity", i.e., near places where image transitions occur (edges, corners, peaks, ...). This property suggests that a spatially adaptive denoising algorithm can be based on selecting the subset that consists of large coefficients and treating it in a different way than the other wavelet coefficients that represent mainly noise. This is the idea of wavelet thresholding [23]-[26]. These techniques were mainly developed for the orthogonal wavelet representation and for the case of additive white Gaussian noise.

Two standard thresholding policies are: *hard-thresholding*, ("keep or kill"), and *soft-thresholding* ("shrink or kill"). In both cases, the coefficients that are below a certain threshold are set to zero. In hard-thresholding, the remaining coefficients are kept unchanged. On the other side, in soft thresholding, the magnitudes of the remaining coefficients are reduced by an amount equal to the value of the threshold. We can say that in both cases each wavelet coefficient is multiplied by a given *shrinkage factor*, which is a function of the magnitude of the coefficient (Fig. 3.1).

In soft thresholding, the estimates are biased: large coefficients are always reduced in magnitude; therefore, the mathematical expectations of their estimates differ from the observed values. On the other side, hard thresholded estimates have a larger variance and may be highly sensitive to small changes in the data.

The choice of an "optimum" threshold depends on the chosen optimality criterion (e.g., the minimization of the expected mean squared error or minimization of a given risk function). Many methods for estimating the threshold use the fact that under orthogonal transformation, such as decomposition into a wavelet basis, the white noise in the input image maps into a white noise in the transformation (wavelet) domain. The well known *universal threshold* [23] is derived as: $T_{univ} = \hat{\sigma}\sqrt{2\log(n)}$, where $\hat{\sigma}$ is the estimate of the standard deviation of additive white noise and n is the total number of the wavelet coefficients in a given detail image. The motivation for choosing

this threshold was the following: if we have a sequence of n independent, identically distributed random variables (white noise) with normal distribution $N(0,1)$, then the maximum of this sequence is smaller than $\sqrt{2\log(n)}$ with a probability approaching one when n tends to infinity. Therefore, for detail images of the same orientation the universal threshold differs only in the constant factor that is related to the number of coefficients at a given resolution level. Other thresholds that are estimated in an adaptive way for each level were proposed, e.g., in [26, 38, 67].

3.2. MAP estimation assuming independent wavelet coefficients. Another approach in wavelet-based image denoising applies classical Bayes estimation theory. Many techniques, like [61, 65], use orthogonal wavelet representation and assume statistically *independent* wavelet coefficients with heavy-tailed prior distributions. We shall explain this approach in more detail, because later (in Section 5) it will be extended to include the spatial dependencies between the wavelet coefficients.

In the estimation theory, the Bayes estimate is known as the best estimate in the case when the data distribution is known and when appreciable prior knowledge about the quantity being estimated is available. A special case of the Bayes estimate is the *maximum a posteriori* (MAP) estimate, which maximizes the *posterior probability* of the solution

$$P(\mathbf{Y}=\mathbf{y}|\mathbf{W}=\mathbf{w}) = \frac{p_{\mathbf{W}|\mathbf{Y}}(\mathbf{w}|\mathbf{y})P(\mathbf{Y}=\mathbf{y})}{p_{\mathbf{W}}(\mathbf{w})}. \tag{3.2}$$

The denominator in (3.2) does not play a role in the search for the vector $\mathbf{y}$ that maximizes the whole expression. Therefore, it can be omitted for the sake of clarity. Under the assumption of an additive white noise model, the conditional density is separable: $p_{\mathbf{W}|\mathbf{Y}}(\mathbf{w}|\mathbf{y}) = \prod_s p_{W_s|Y_s}(w_s|y_s)$. Further on, if it is assumed that the wavelet coefficients are independent, the prior distribution is also separable, i.e., $P(\mathbf{Y}=\mathbf{y}) = \prod_s P(Y_s = y_s)$. In this case, the MAP estimation problem can be separated into n independent optimization problems:

$$\hat{y}_s = \arg\max_s \, p_{W_s|Y_s}(w_s|y_s)P(Y_s=y_s), \qquad 1<s<n. \tag{3.3}$$

The prior distribution of the wavelet coefficients in the form of a generalized Laplacian distribution $p_Y(y) = a\exp(-|by|^\nu)$ was proposed in [47], where the *shape parameter* ν is typically between 0 and 0.5. Estimation of the parameters of the prior distribution using higher-order statistics was performed in [61]. The relationship between the MAP estimators and wavelet thresholding was studied in [53]. There it was shown that the soft-thresholding method is equivalent to the MAP estimation assuming a Laplacian prior on the wavelet coefficients, with standard deviation equal to $\sigma_y = \sigma^2 T_{univ}^{-1}\sqrt{2}$, where σ^2 is the noise variance and T_{univ} is the universal threshold.

3.3. The use of a redundant wavelet representation in denoising. A redundant wavelet representation offers a better reconstruction of useful signal from noise than the orthogonal one (see, e.g., the discussion in [16]). We assume here the specific redundant representation with the same number of wavelet coefficients at each resolution scale. In this case, as it can be seen from Fig. 2.3, the most significant wavelet coefficients are highly correlated across scales and inside a given detail image. Different denoising techniques make use of these dependencies in different ways, in order to achieve a better separation of useful signal from noise. In [48], the evolution of the modulus maxima of the wavelet transform through resolution scales is exploited to estimate the local Lipschitz exponents α (see Section 2.1). The modulus maxima

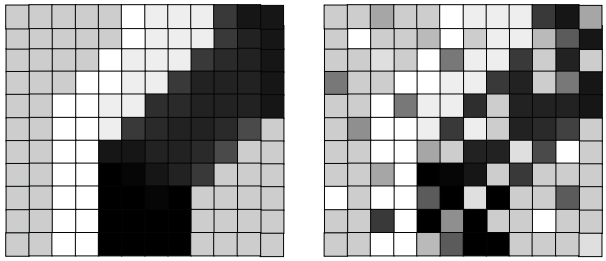


FIGURE 4.1. *(a) An example of spatially connected clusters of wavelet coefficients. (b) A perturbation of (a) possibly due to noise.*

which originate from discontinuities with $\alpha < 0$ are assumed to represent noise and are removed. The denoised image is then reconstructed from these modified modulus maxima. It was theoretically proved and practically demonstrated [49] that a close approximation of the image can be reconstructed from the modulus maxima of the wavelet transform. However, the reconstruction involves a computationally demanding iterative procedure. Several techniques, like [11], have been proposed to improve the convergence of the reconstruction process. The approach of [36] performs the thresholding of the wavelet coefficients, where the threshold is applied to the rough estimates of the local Lipschitz exponents; these estimates are found by computing the interscale ratios of the wavelet coefficients. Another method [66], calculates the products between the wavelet coefficients at several adjacent scales in order to identify the positions of significant wavelet coefficients, from which the denoised signal is reconstructed.

4. Markov Random Field Image Models. The wavelet based denoising methods analyzed so far do not exploit *spatial dependencies* between wavelet coefficients. Generally, the use of spatial context is of great importance in image processing and pattern recognition. The intensity, i.e., the brightness level of a pixel in an image is highly dependent on the intensities of surrounding pixels, unless the image is simply random noise. The knowledge about the image structure can be used to recover missing information or to correct erroneous data. The image structure is also reflected in the detail images of its wavelet decomposition. We assume here the redundant decomposition, with equal number of the wavelet coefficients at each scale (Fig. 2.3). In this case, high spatial dependencies between wavelet coefficients are evident.

An example in Fig. 4.1(a) illustrates schematically a part of a typical detail image. The lightest colors correspond to the largest positive coefficients, and the darkest colors to the negative coefficients with highest magnitudes. The spatially connected clusters, like these in Fig. 4.1(a), are *a priori* more probable, than the case where the large coefficients appear isolated like in Fig. 4.1(b). This prior knowledge can be used to improve the result of the denoising. The dependencies among neighboring pixel intensities are quantitatively expressed by using an appropriate *image model*. There exist many different image models, such as linear prediction [37], random mosaic [2], fractal models [30, 55], etc. Random field image models treat the intensity of a pixel as a random variable and impose statistical dependence between image pixels.

It is convenient to introduce at this point the notion of a *label* as it will be frequently used in the rest of this chapter. Many problems in image processing, such as image restoration, segmentation or edge detection fit into a general image labeling

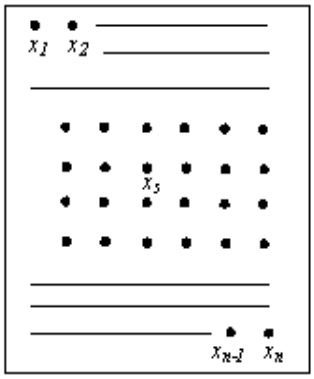


FIGURE 4.2. *A configuration (realization) of an image field sampled on a rectangular lattice.*

framework, where a given label is assigned to each pixel. In image restoration, the label that is assigned to a pixel is the "true" gray value of its intensity. In image segmentation the label determines the category of a pixel, i.e., its belonging to a specific class. In edge detection the labels are binary: if a pixel belongs to an edge it is assigned one label and in the opposite case the other. The values of the wavelet coefficients, in this general framework, can also be interpreted as labels assigned to the corresponding pixels in detail images. In the language of random image models the statistical dependencies among pixel labels are called *spatial interactions*.

4.1. Markov Random Fields. MRF models are widely used in image processing and pattern recognition, as they provide a convenient way of modeling *local* spatial interactions, i.e., they describe statistical dependence of a pixel label on labels in its local spatial surrounding. In real images, a pixel label (e.g., the gray value in a given point) is significantly correlated only with a limited number of other labels and for this reason the local characterization is valid. For image denoising the *global* characterization in terms of the joint probability of all pixel labels is also important, but this subject will be addressed later.

First, we introduce the notations and some basic concepts of random image models. Let $S = \{1, ..., n\}$ be a finite index set - the set of *sites* on a regular rectangular lattice. The elements of S correspond to points at which an image is sampled, i.e., to the location of image pixels. A family of random variables $\mathbf{X} = \{X_1, ..., X_n\}$ defined on the set S is called the *image field*. The notation $\mathbf{X} = \mathbf{x}$ will be used to abbreviate the joint event $(X_1 = x_1, ..., X_n = x_n)$. The vector $\mathbf{x} = \{x_1, ..., x_n\}$ is a *configuration* of $\mathbf{X}$, corresponding to a given realization of the image field. The space of all possible configurations of $\mathbf{X}$ will be denoted by $\mathcal{X}$. A subscript in the notation of a vector will be used to indicate that only some variables are present in the vector. For example, $\mathbf{X}_{S\setminus s}$ denotes the vector of random variables X_t such that $t \in S \setminus \{s\}$.

A *random field* is a family of random variables $\mathbf{X} = \{X_1, ..., X_n\}$ such that all its possible configurations have strictly positive probability. A specific class of random field models called Markov random fields furthermore requires that the label of each pixel is influenced only by pixels that are its *neighbors*. Not necessarily, but usually these neighbors are the pixels that are surrounding the current one. Formally, the neighboring relation is defined as follows.

DEFINITION 4.1. *A collection $\partial = \{\partial(s) : s \in S\}$ of subsets of S is called a* **neighborhood system**, *if*

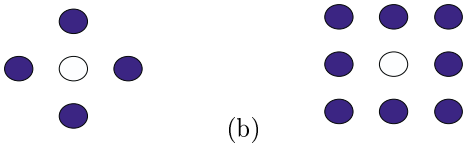


FIGURE 4.3. *(a) First order neighborhood. (b) Second order neighborhood.*

1. $s \notin \partial(s)$
2. $s \in \partial(t)$ *if and only if* $t \in \partial(s)$.

The sites $s \in \partial(t)$ *are called* **neighbors** *of* t. *The notation* $\langle s, t \rangle$ *will be used to denote that pixels* s *and* t *are neighbors of one another.*

Two examples of neighborhoods are shown in Fig. 4.3. The neighborhood in Fig. 4.3(a) is called the four-point or the *first order* neighborhood, and the neighborhood in Fig. 4.3(b) is called the eight-point or the *second order* neighborhood. For practical reasons, these two neighborhood systems are the most frequently used in image processing. The formal definition of a Markov random field is:

DEFINITION 4.2. *The* random *field* $\mathbf{X}$ *is a* **Markov field** *with respect to the neighborhood system* ∂ *if for all* $\mathbf{x} \in \mathcal{X}$

$$P(X_s = x_s | \mathbf{X}_{S \setminus s} = \mathbf{x}_{S \setminus s}) = P(X_s = x_s | \mathbf{X}_{\partial(s)} = \mathbf{x}_{\partial(s)}). \tag{4.1}$$

According to this definition, the probability of a pixel label x_s, given all other labels in the image practically reduces to a function of relatively small number of labels that are in the neighborhood $\partial(s)$. It should be noted that by choosing an arbitrarily large neighborhood $\partial(s)$ the MRF model can be applied to every image. MRF models that are applied in image processing are often *homogeneous* (i.e., strictly stationary), but it is not a general property of a MRF. The homogeneity property means that the distribution $P(X_s = x_s | \mathbf{X}_{\partial(s)} = \mathbf{x}_{\partial(s)})$ is the same for all pixels s.

4.2. The equivalence between Markov and Gibbs random fields. Gibbs random fields (GRF) [52] are used in statistical mechanics as probability models for the fluctuations of large physical systems around their equilibrium state. While MRFs characterize local aspects, GRFs explicitly express the joint probability of the system variables, providing thus a model for *global context*. The link between Markov and Gibbs random fields provides an excellent framework for specifying the global context via local spatial interactions. A Gibbs random field is generally a random field $\mathbf{X}$ for which the configurations $\mathbf{x}$ obey a Gibbs distribution

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z}\, e^{-H(\mathbf{x})/T}, \tag{4.2}$$

where $H(\mathbf{x})$ is called the *energy function* and Z and T are constants. Z is the normalizing constant $Z = \sum_{\mathbf{x} \in \mathcal{X}} e^{-H(\mathbf{x})/T}$ and is called the *partition function*. T is the *temperature* and it actually controls the "peaking" in the probability density: if T is smaller there is a bigger difference between joint probabilities of different configurations $\mathbf{x}$ and peaks in the density that correspond to configurations with lowest energy are easier to find.

The relationship between Gibbs and Markov random fields was established by introducing cliques (Fig. 4.4). A *clique* is a set of sites in which two different elements are always neighbors. The set of all cliques will be denoted by $\mathcal{C}$. The *clique potential* $V_C(\mathbf{x})$ is a function of only those variables x_s for which $s \in C$. If the energy $H(\mathbf{x})$ can

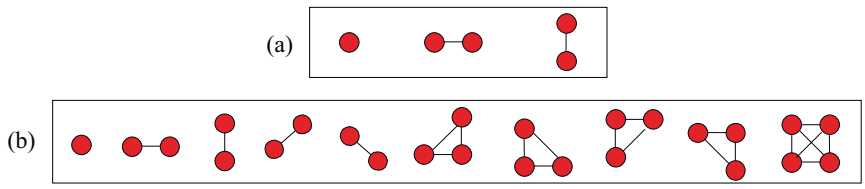


FIGURE 4.4. *(a) The clique types for (a) the first order and (b) the second order neighborhood.*

be decomposed as a sum of clique potentials then the corresponding Gibbs random field is called a neighbor Gibbs field, and is equivalent to a Markov random field. This is stated in the often cited Hammersley-Clifford theorem:

THEOREM 4.3. *A random field is a Markov field for the neighborhood system ∂ if and only if it is a neighbor Gibbs field for ∂.*

Different proofs of this theorem can be found in, e.g., [44, 68]. Its practical value is that it provides a simple way of specifying the joint probability of MRFs:

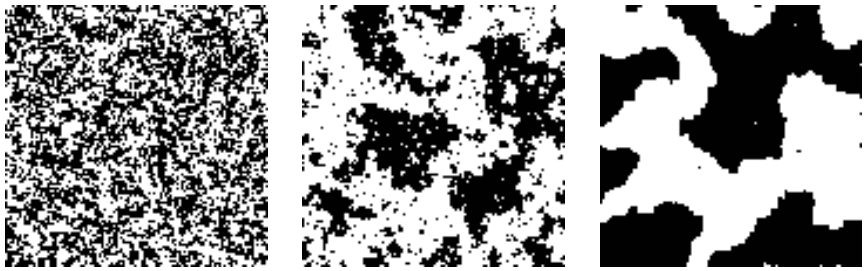
$$P(\mathbf{X}=\mathbf{x}) = \frac{1}{Z}\exp\Big(-\frac{1}{T}\sum_{C\in\mathcal{C}} V_C(\mathbf{x})\Big). \tag{4.3}$$

The equation (4.3) characterizes the *global* context (joint distribution of all pixels in an image) in terms of *local* spatial characteristics that are expressed through clique potentials. One chooses the appropriate clique potential functions to give preference to certain local spatial interactions. In this way, the *prior* knowledge about an image is encoded. For example, for image denoising it is important to remove the occurrence of isolated pixels that are usually produced by noise. This prior knowledge can be encoded by choosing such a clique potential function that the lowest potential corresponds to the case when all pixel labels in a clique are equal. In this way we give a preference to spatial continuity over pixels that differ strongly from their neighbors.

4.3. Specification of MRF models. There exist different MRF models that assume continuous or discrete label sets and that use different neighborhoods and clique potential functions. The particularly simple *Ising* model is often used as a starting point for studying Markov Random Fields. This model was proposed by the German physicist E. Ising in 1925 for the purpose of modeling the behaviour of ferromagnets. We shall present it here in terms of images. The energy function in the Ising model takes the form

$$H(\mathbf{x}) = \alpha\sum_s x_s + \beta\sum_{\langle s,t\rangle} x_s x_t, \tag{4.4}$$

where α and β are constants and the labels x_s assume binary values ± 1. The neighbors $\langle s,t\rangle$ are the horizontally and vertically adjacent pixels at the Euclidean distance 1; thus the first order neighborhood from Fig. 4.3 is assumed. The first term in (4.4) is the *cost* that is paid for the occurrence of a given label in the image irrespective of the values of its neighbors: depending on the sign of the factor α one of the labels -1 or +1 will contribute to the higher energy, being thus "more expensive" and less probable. Therefore, by choosing α we encode the prior preference of a given label. If both labels are equally probable a priori, we choose $\alpha = 0$. The second term in (4.4) corresponds to the interaction of neighboring labels. If $\beta < 0$ then equal neighboring



(a) (b) (c)

FIGURE 4.5. *Different configurations of the Ising model. In all cases* $\alpha = 0$ *and (a)* $\beta = -0.25$, *(b)* $\beta = -0.45$ *and (c)* $\beta = -0.9$.

labels contribute to the lower energy, and therefore $x_s = x_t$ will be more probable than $x_s \neq x_t$. If $\beta > 0$ the opposite is true. The Ising model is *homogeneous* because α and β are constants that do not depend on the spatial location s. It is also *isotropic* (i.e., rotationally invariant) because β is the same for neighbors $\langle s, t\rangle$ in all possible (here horizontal and vertical) directions. To illustrate the Ising model, we present in Fig. 4.5 three samples of this random field that correspond to $\alpha = 0$ and different values of the parameter β. We have sampled these configurations using the Metropolis algorithm (Section 5.3) at temperature $T = 1$.

For practical reasons, image processing methods mostly use MRF models with cliques consisting of two sites (*pair-site* cliques). When the prior probabilities of different labels are known (which is less frequently the case in practice), single-site cliques are additionally used. The general form of the energy for MRF models with clique potentials of up to two sites is

$$H(\mathbf{x}) = \sum_{C \in \mathcal{C}} V_C(\mathbf{x}) = \sum_{\langle s,t\rangle} V_2(x_s, x_t) + \sum_s V_1(x_s), \tag{4.5}$$

where $V_1(x_s)$ denotes the potential function for single-site cliques, and $V_2(x_s, x_t)$ is the potential function for pair-site cliques.

Commonly used MRF models in image processing are *automodels* [5]. There, the potential functions are of the form $V_1(x_s) = x_s g_s(x_s)$ and $V_2(x_s, x_t) = \beta_{s,t} x_s x_t$; $g_s(\cdot)$ are arbitrary functions and $\beta_{s,t}$ are constants, which reflect the pair-site interactions. If the constants $\beta_{s,t}$ are different for pairs $\langle s, t\rangle$ that form cliques of different orientations then preference is given to spatial clusters in specific directions; this is needed for modeling anisotropic textures [19]. For image denoising, isotropic models are mainly used. However, if we apply a MRF model to encode the prior knowledge about wavelet detail images, it could be useful to use an anisotropic model to give preference to horizontal, vertical or diagonal clusters in different detail images.

Important classes of auto models are *auto-logistic* models [44], *auto-binomial* models [19] and *auto-normal* models, which are also called Gaussian Markov Random Field (GMRF) models [13, 14, 39]. GMRF models are often used for image segmentation [8, 17, 43]. In [41], a method was proposed for selecting the neighbor sets in GMRF, based on the spatial correlation of image.

The *multi-level logistic* (MLL) model is used, e.g., in [22] and in [32], where it is called the *generalized Ising* model. For cliques containing two or more sites the MLL

potentials are defined as

$$V_C(\mathbf{x}) = \begin{cases} -\gamma_c, & \text{if all } x_s, s \in C \text{ are equal,} \\ +\gamma_c, & \text{otherwise,} \end{cases} \tag{4.6}$$

where γ_c is the potential for type c cliques. We have used this model, with pair-site cliques and the second order neighborhood for the results in Section 6.

5. MAP-MRF approach to wavelet based image denoising. In this section, we address the MAP approach to wavelet based image denoising, applying a MRF prior model. We use the redundant wavelet representation, with equal number of the wavelet coefficients at each resolution scale (Section 2.3). The assumed noise model is given in Eq (3.1). However, the analysis that follows can be applied or easily extended to other cases, with different kinds of noise.

5.1. Bayesian estimation of the wavelet coefficients using a MRF prior. The MAP approach was already introduced in Section 3.2. However, there it was assumed that a prior distribution of each wavelet coefficient was available and that wavelet coefficients were statistically independent. Now that we take the spatial dependencies into account, we search for the joint MAP estimate $\hat{\mathbf{y}}$ of all wavelet coefficients in a detail image

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}\in\mathcal{Y}} p_{\mathbf{W}|\mathbf{Y}}(\mathbf{w}|\mathbf{y}) P(\mathbf{Y}=\mathbf{y}), \tag{5.1}$$

where the joint probability $P(\mathbf{Y}=\mathbf{y})$ is a Gibbs distribution (4.3). Similarly as in Section 3.2, we assume $p_{\mathbf{W}|\mathbf{Y}}(\mathbf{w}|\mathbf{y}) = \prod_s p_{W_s|Y_s}(w_s|y_s)$. This is a reasonable assumption when noise is spatially noncorrelated. Furthermore, one can present an arbitrary conditional distribution in the exponential form $p(w|y) = A\exp(-V(w|y))$ by choosing $V(w|y) = \ln(A) - \ln(p(w|y))$, where A is a given constant. Then the conditional probability model becomes

$$p_{\mathbf{W}|\mathbf{Y}}(\mathbf{w}|\mathbf{y}) = A\exp\Big(-\sum_s V(w_s|y_s)\Big). \tag{5.2}$$

If we use this conditional probability model and if we use the Gibbs distribution in (4.3) with the partition function Z, the temperature T and the energy function $H(\mathbf{y}) = \sum_{C\in C} V_C(\mathbf{y})$ to model the joint probability $P(\mathbf{Y}=\mathbf{y})$ of the prior MRF field, the posterior probability is also a Gibbs distribution

$$P(\mathbf{Y}=\mathbf{y}|\mathbf{W}=\mathbf{w}) = \frac{1}{\zeta}\ \exp\Big(-\frac{H(\mathbf{y}|\mathbf{w})}{\tau}\Big), \tag{5.3}$$

where the partition function is $\zeta = A/(Z\cdot P(\mathbf{W}=\mathbf{w}))$, which is independent of $\mathbf{y}$. The temperature is $\tau = T$ and the *posterior energy* is of the form

$$H(\mathbf{y}|\mathbf{w}) = \sum_s V(w_s|y_s) + \lambda \sum_{C\in\mathcal{C}} V_C(\mathbf{y}). \tag{5.4}$$

The constant λ actually represents the influence of the prior knowledge that is encoded in the model. Thus, the MAP solution (5.1) is the global minimum of the posterior energy (5.4). This is a classical image restoration problem using Gibbs distributions [32], applied to denoising in the wavelet domain. The posterior probability (5.3) is

an imaginary physical system whose lowest energy state is exactly the MAP estimate of the noise-free detail image given the noisy data. In the particular case where the observation noise is white Gaussian and has zero mean and variance σ^2, the posterior energy is of the form

$$H(\mathbf{y}|\mathbf{w}) = \sum_s (w_s - y_s)^2 + \frac{2\sigma^2}{T} \sum_{C \in \mathcal{C}} V_C(\mathbf{y}). \tag{5.5}$$

Practical computation of the MAP solution $\hat{\mathbf{y}}$ by minimizing the posterior energy $H(\mathbf{y}|\mathbf{w})$ is addressed in Section 5.3. Such a solution should yield better results than one assuming independent wavelet coefficients, if the clique potentials are correctly defined.

The specification of clique potentials is a key problem: they should penalize the abrupt changes due to noise, but not the actual discontinuities. This problem is in general important in image denoising using MRF priors. In classical, single-resolution techniques some approaches involve "line processes", which are updated in parallel with the Markov random field [31, 32]. Many different "discontinuity adaptive" (DA) methods [44] have been proposed. Generally speaking, DA potentials are nonlinear functions of the absolute difference between neighboring pixel intensities, which tend to turn off the smoothing when this difference is large. The wavelet representation naturally facilitates various discontinuity adaptive approaches: large wavelet coefficients occur exclusively at the positions of image discontinuities. Moreover, if these discontinuities are produced by noise the magnitudes of the corresponding coefficients rapidly decrease through scales. Due to these properties, the positions of meaningful edges are more reliably detected than in the single-resolution case. The construction of different edge adaptive potentials in this framework is an interesting problem, which is not well studied yet. For example, the energy of the wavelet coefficients in local surroundings can be used to indicate the presence of an edge [28]. Such a local measurement can be introduced as an additional variable in clique potentials. Similarly, the measurement derived from several resolution scales, at a given spatial position, may be used to control the clique potential.

Another problem with the MAP-MRF approach is the computation time (see Section 5.3). It was proved [8, 33] that better results can be obtained, and the computation time reduced by applying a multiscale approach and a coarse-to-fine strategy. This processing combines optimization at different resolution scales with an *interscale* transfer of information. The optimization criterion is the same in this case: the minimization of the posterior energy at the corresponding scale. The interscale information can be introduced in different ways. For example, the solution from the previously processed scale is the initial solution for the current scale. Another possibility is to consider each scale as causally dependent on the previously processed one, and to introduce the corresponding interscale cliques. This approach can be efficiently applied in the wavelet domain.

5.2. Bayesian shrinkage of the wavelet coefficients using a MRF prior. The simple shrinkage of *individual* wavelet coefficients (Section 3) provides reasonably good denoising results in many practical cases. To improve the heuristic shrinkage estimators leading to the choice of a given threshold, Bayesian "mimicking" of thresholding rules was used in [65]; it takes into account the prior distribution of individual wavelet coefficients. Further improvements can be obtained by taking into account spatial dependencies between wavelet coefficients as well. Such an approach was first proposed in [51], and later extended in other papers, e.g., [58].

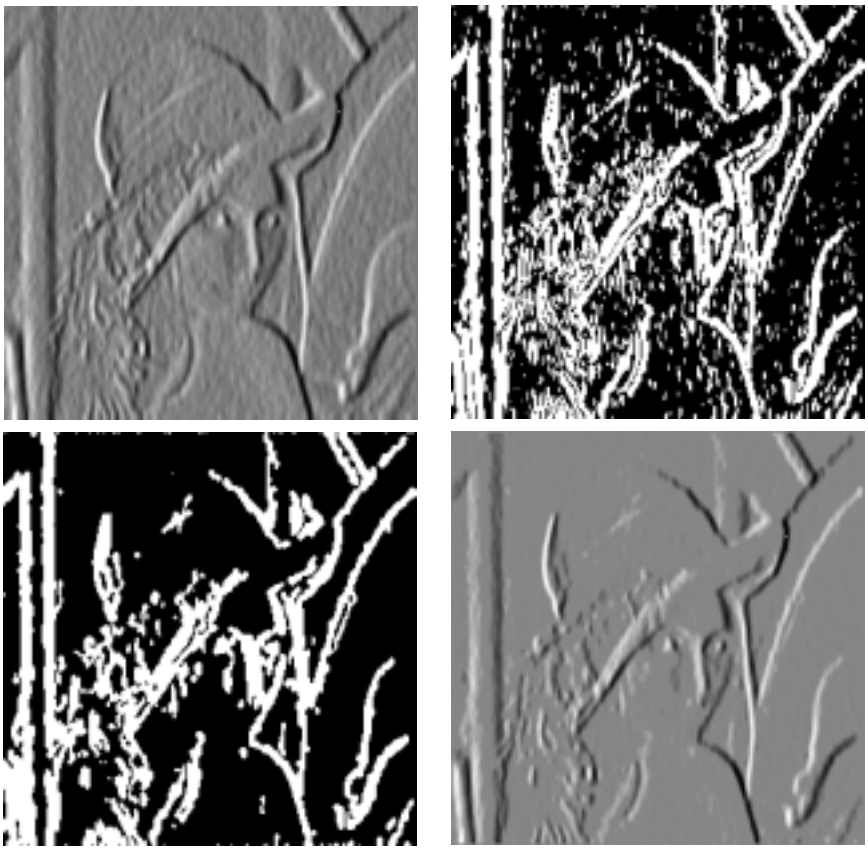


FIGURE 5.1. *Top left: a noisy detail image. Top right: an initial classification of binary labels. Bottom left: the MAP estimate of the mask. Bottom right: cleaned detail.*

The basic concept is the following: for each detail image $\mathbf{w}$ we estimate a *mask* $\mathbf{x}$, which indicates the positions of meaningful edges. Subsequently, an edge-adaptive suppression is applied to the wavelet coefficients, using the estimated mask. In Fig. 5.1, we illustrate this approach. The mask $\mathbf{x}$ is a set $\{x_1, ..., x_n\}$ of binary labels: $x_s = 0$ if w_s is assumed to originate from noise and $x_s = 1$ if w_s is assumed to represent a useful signal. To estimate masks, we define for each wavelet coefficient w_s a given measure of its significance m_s, and we exploit the prior knowledge about the spatial clustering of labels x_s. Spatial clustering is encoded using a MRF prior model. In other words, the vector $\mathbf{x}$ is assumed to be a specific configuration of a Markov random field $\mathbf{X}$. The measure of significance m_s is computed from the observed wavelet coefficients. For example, one can define m_s as the magnitude of the wavelet coefficient $|w_s|$, an estimate of the local Lipschitz exponent, or the amount of interscale correlation at the position s. In these cases a larger value of m_s indicates that w_s is more likely to represent a useful signal, while a smaller value of m_s indicates that w_s is more likely to originate from noise. Using such reasoning, a heuristic model for the likelihood $p_{M_s|X_s}(m_s|x_s)$ was proposed in [51]. In [58], we have proposed a more accurate statistical characterization of different significance measures, which yields realistic models for $p_{M_s|X_s}(m_s|x_s)$. Similarly as in Section 5.2, we can represent the

conditional probability density in the form

$$p_{\mathbf{M}|\mathbf{X}}(\mathbf{m}|\mathbf{x}) = A \exp\Big(-\sum_s V(m_s|x_s)\Big), \tag{5.6}$$

where A is an arbitrary constant. The MAP estimate of the mask $\mathbf{x}$ is found by minimizing the posterior energy

$$H(\mathbf{x}|\mathbf{m}) = \sum_s V(m_s|x_s) + \lambda \sum_s V_C(\mathbf{x}), \tag{5.7}$$

where the constant λ like in (5.4) controls the influence of the prior knowledge about spatial clustering. This approach offers an elegant way of combining inter- and intrascale dependencies between wavelet coefficients. For example, if one defines m_s as the interscale correlation or the averaged interscale ratio at spatial position s, then the interscale dependencies are captured in the data distribution model $V(m_s|x_s)$. Intra-scale dependencies are modeled via spatial clustering of binary labels.

One can use the estimated binary mask $\hat{\mathbf{x}}$ in different ways to modify the wavelet coefficients. The simplest approach is the hard thresholding

$$\hat{y}_s = \hat{x}_s w_s. \tag{5.8}$$

This simple manipulation of wavelet coefficients already yields better results than the classical thresholding according to their magnitude alone. The labels $\hat{x}_s$ were estimated by taking into account not only the local measure (like magnitude of the coefficient) but also the spatial interactions between the wavelet coefficients. Even better results can be obtained if the hard binary decision in (5.8) is replaced by a soft modification rule, as was done in [51]:

$$\hat{y}_s = P(X_s = 1|\mathbf{M}) w_s. \tag{5.9}$$

This produces a soft shrinkage of the wavelet coefficients: those coefficients that are more likely to originate from noise are attenuated more than those coefficients that are likely to represent a useful signal. To compute the marginal probabilities $P(X_s = 1|M)$, strictly according to the theory, one needs to know the joint posterior probability $P(\mathbf{X}|\mathbf{M})$. In practice *stochastic samplers* like the Metropolis algorithm (Section 5.3) are used to obtain reliable estimates: the probability $P(X_s = 1|M)$ is estimated by measuring the frequency of occurrence of the label value 1 at spatial position s, in the chain of masks generated by a random sampling algorithm.

We have proposed in [56] another, much faster approach for the soft modification of the wavelet coefficients, taking into account spatial interactions. There, the estimates of the wavelet coefficients are found as

$$\hat{y}_s = q(m_s, u_s) w_s, \quad u(s) = \sum_{t \in \partial(s)} (2\hat{x}_t - 1) \tag{5.10}$$

where the function $q(m_s, u_s)$ is a nonlinear function, bounded between zero and one, and u_s is the neighborhood parameter, which is computed from the MAP estimate of the mask. The shrinkage factor $q(m_s, u_s)$ is monotonically increasing in a given interval around the specified threshold T_{hr}:

$$q(m_s, u_s) = \frac{\xi(m_s)^\alpha \ \eta(u_s)^\beta}{1 \ + \ \xi(m_s)^\alpha \ \eta(u_s)^\beta}, \qquad (1-\delta)T_{hr} < m_s < (1+\delta)T_{hr}; \tag{5.11}$$

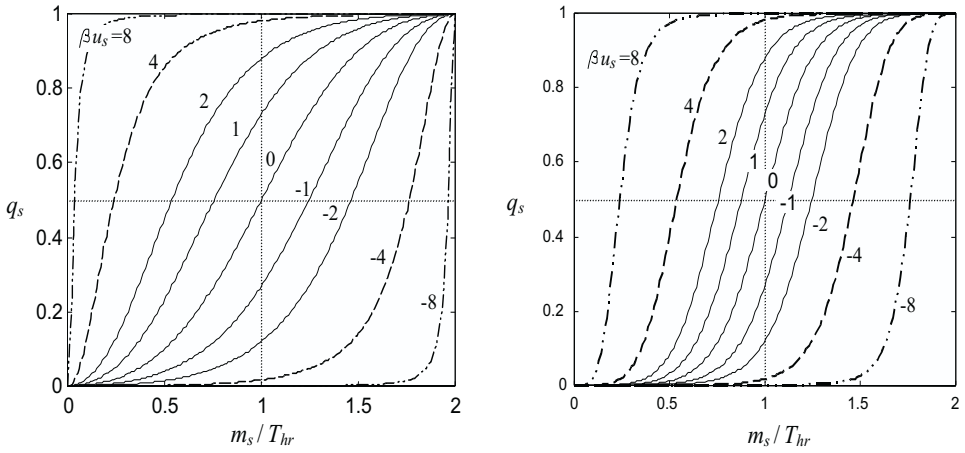


FIGURE 5.2. *The shrinkage factor $q_s = q(m_s, u_s)$ versus normalized measure m_s/T_{hr}. The parameter is the product βu_s. (a) $\delta = 1$, $\alpha = 2$ and (b) $\delta = 1$, $\alpha = 4$.*

outside of this interval, $q(m_s, u_s) = 0$ for $m_s < (1-\delta)T$ and $q(m_s, u_s) = 1$ for $m_s > (1+\delta)T$. The parameters $\alpha \geq 0$ and $\beta \geq 0$ control the influence of the local measures and the spatial interactions, respectively, and

$$\xi(m_s) = \frac{m_s - (1-\delta)T_{hr}}{(1+\delta)T_{hr} - m_s}, \qquad \eta(u_s) = \exp(u_s). \tag{5.12}$$

The shrinkage factor in (5.10) - (5.12) satisfies the following conditions. If the neighborhood $\partial(s)$ is "neutral" (an equal number of neighboring labels have values 0 and 1), or if we do not give any influence to the prior knowledge ($\beta = 0$), then the shrinkage factor is influenced only by m_s. In this case, the shrinkage function can vary between a linear function (for $\alpha = 1$) and the hard thresholding (for $\alpha \to \infty$). In case where $\beta > 0$, the shrinkage factor is closer to one when more labels in $\partial(s)$ indicate an edge. The greater β is, the greater is the influence of the neighboring labels. Another, interpretation of the shrinkage function defined by (5.10) - (5.12) is the following. The marginal probability $P(X_s = 1 | \mathbf{M} = \mathbf{m}, \mathbf{X}_{S\setminus s} = \hat{\mathbf{x}}_{S\setminus s})$ that the wavelet coefficient w_s is useful, given the set of significance measures $\mathbf{m}$ and given the estimated mask at all the positions except the current one $\hat{\mathbf{x}}_{S\setminus s}$, can be represented [57] in the form $\xi_s\eta_s/(1+\xi_s\eta_s)$, where ξ_s is the ratio of likelihood functions $\xi_s = p_{M_s|X_s}(m_s|1)/p_{M_s|X_s}(m_s|0)$, and η_s is the ratio of prior probabilities $\eta_s = P(X_s = 1|\hat{\mathbf{x}}_{\partial(s)})/P(X_s = 0|\hat{\mathbf{x}}_{\partial(s)})$. The ratio of likelihood functions ξ_s becomes exactly equal to $\xi(m_s)$ from (5.12) if $p_{M_s|X_s}(m_s|x_s)$ are approximated by simple piecewise linear functions, with two constant parts and a linear transition around the threshold. For the MRF prior model with pair-site cliques and the potentials (4.6), the ratio of prior probabilities η_s is exactly equal to $\eta(m_s)$ from (5.12).

Two different families of characteristics $q(m_s, u_s)$ are shown in Fig. 5.2. It can be seen that if the significance measure of the coefficient is equal to the chosen threshold ($m_s/T_{hr} = 1$) and if the neighborhood $\partial(s)$ contains the same number of edge and non-edge labels ($u_s = 0$), then the shrinkage factor $q(m_s, u_s)$ is equal to 0.5. Depending on its neighborhood, a coefficient can be significantly reduced even if its significance measure is highly above the threshold. On the other hand, if the neighborhood indicates an edge the coefficient is less suppressed.

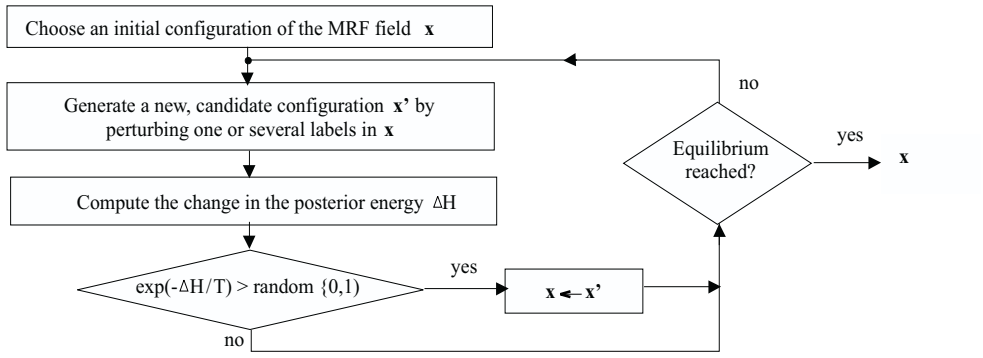


FIGURE 5.3. *The Metropolis algorithm at the temperature T.*

5.3. Computational aspects. For both problems that were addressed in this Section, the MAP estimation of the detail images and the MAP estimation of masks, we need to minimize the posterior energies of the same form. The significant difference is that in the second case we operate with binary labels, which is an easier computational problem. If the values of the wavelet coefficients are quantized, the same combinatorial minimization techniques can be used for both problems.

The minimization of the energy function by deterministic gradient descent algorithms is likely to end up in a local minimum. To overcome this problem, random search methods are often used, such as the *Metropolis algorithm* [44] and the *Gibbs sampler* [32]. These algorithms allow occasional increases of the energy, in order to get out of a local minimum. Random search starts from some initial estimate and a chain of configurations is generated by randomly perturbing labels. The convergence is faster if the initial estimate is good. For masks, an initial estimate is simply obtained by using a given threshold for the chosen significance measure of the wavelet coefficients. The Metropolis algorithm is schematically depicted in Fig. 5.3. During the random search process, a new "candidate" configuration is obtained by random perturbation of the previous configuration; subsequently, it is accepted or not, depending on the change in the posterior energy ΔH. If $\Delta H \leq 0$ the change is accepted and if $\Delta H > 0$ the change is accepted with probability p. Practically, a random number with uniform distribution on [0,1) is generated and compared with $\exp(-\Delta H/T)$. When all the labels are updated, one iteration is completed. This method does not guarantee the global minimum solution, but a low energy configuration is found with a large probability. In general, a sufficient number of generated sample configurations to reach the equilibrium is 50 times the size of the image [44]. For masks of binary labels, that we consider, 10 iterations usually suffice if the initial estimate is good.

In image processing applications, the *iterated conditional modes* (ICM) technique [6] is often used. It searches for a solution close to the MAP one, by maximizing the local conditional probabilities $P(X_s = x_s | \mathbf{Y} = \mathbf{y}, \mathbf{X}_{S\setminus s} = \mathbf{x}_{S\setminus s})$ sequentially. For the serial updating case, convergence is guaranteed and is rapid (5 to 6 raster scans of an image are sufficient). For masks in our application, this method is very efficient. It provides similar estimates as the Metropolis algorithm, while the number of iterations is at least halved.

Other optimization techniques, like random and deterministic *annealing* algorithms, are discussed in [44]. Annealing algorithms tend to overcome the problem of

FIGURE 6.1. *Top left: noise-free image. Top right: image with additive white Gaussian noise. Bottom: two results of the same wavelet shrinkage function, without using spatial prior (Left) and using spatial prior (Right).*

getting trapped in local minima by decreasing the value of a given parameter (e.g., temperature in the Gibbs distribution) during the iterative minimization. For the MAP estimation of detail images, in the continuous case, deterministic annealing algorithms, like *graduated non-convexity* (GNC) [7] may be an attractive choice.

6. Applications. Here we present several applications of the Bayesian wavelet shrinkage using MRF priors. First, the advantage of using the estimated masks to achieve a spatially adaptive wavelet shrinkage is demonstrated in Fig. 6.1. The input image is corrupted by additive white Gaussian noise with standard deviation 14. Both presented results were obtained by applying the same shrinkage function (5.11) with $\alpha = 1$, but in one case the information from masks was not used ($\beta = 0$) and in the second case masks were used to control the shrinkage factor ($\beta = 0.2$). In other words, in the first case the shrinkage factor is a function of the chosen significance measure only, like in classical thresholding techniques; in the second case it is extended to a family of functions, depending on the spatial context as well. We

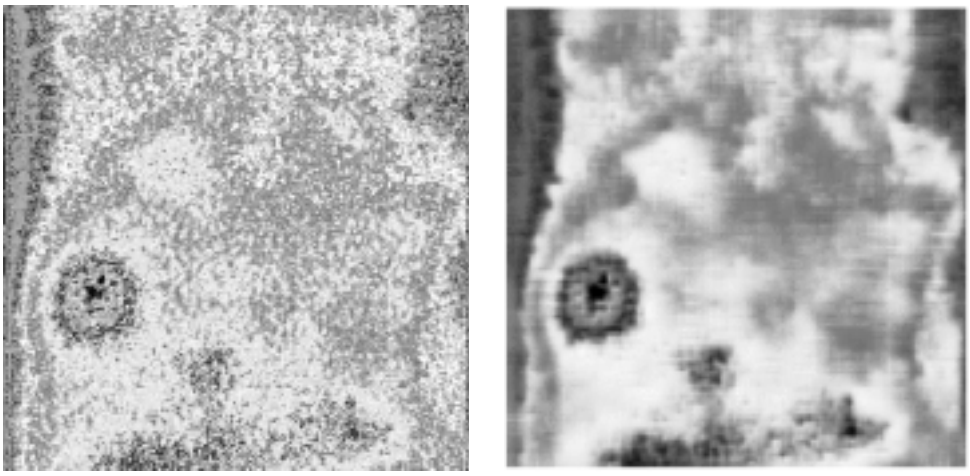


FIGURE 6.2. *Left: an original infrared image. Right: the result of the Bayesian wavelet shrinkage using the MAP-MRF approach.*

have used the second-order neighborhood $\partial(s)$, and the measure of significance m_s was the magnitude of the coefficient. For conditional densities $p_{M_s|X_s}(m_s|x_s)$ we used a simple approximation (similar to one in [51]): piecewise linear, with two constant parts and a linear transition around the chosen threshold. The threshold T_{hr} was chosen as the standard deviation of noise in a given detail image, and the parameter $\delta = 0.5$. Masks were estimated using ICM. By comparing the two results in Fig. 6.1 one can clearly see the advantage of using these masks: noise in background areas is better suppressed while the significant edges are preserved.

Our main interest here is to investigate the applicability of this method in practical cases, where images are contaminated by different kinds of natural noise. To achieve an optimum performance in this respect one needs to examine different choices for the significance measure m_s, and to use realistic likelihood models $p_{M_s|X_s}(m_s|x_s)$. The simulation method that we have proposed in [58] can be helpful in this respect. Here we used in all cases the simple, above described approximation for $p_{M_s|X_s}(m_s|x_s)$. The significance measure was magnitude of the wavelet coefficient in the first application, and a rough estimate of the local Lipschitz exponent in the other two.

The first application that we consider is the denoising of infrared (IR) images in humanitarian demining. In images acquired by infrared sensors, noise is a mixture of thermal (Gaussian) and photonic (Poisson) noise. One IR image of a buried landmine is shown in Fig. 6.2. One can see that Bayesian shrinkage using MRF priors achieves an excellent result in this case. Noise is completely suppressed, while the image preserves the natural variation of intensities, which corresponds to slowly varying temperature of the soil. Details of the landmine are well preserved.

Next, we investigate the applicability of the same approach in remote sensing. One of the most frequently used sensors there is Synthetic Aperture Radar [9], due to its high spatial resolution. SAR images are corrupted by speckle noise [34]. As it can be seen from Fig. 6.3, our denoising method efficiently suppresses speckle noise in SAR images. At the same time, fine structures and the sharpness of edges are preserved well. The processed images are much more convenient for some further image processing tasks, like segmentation.

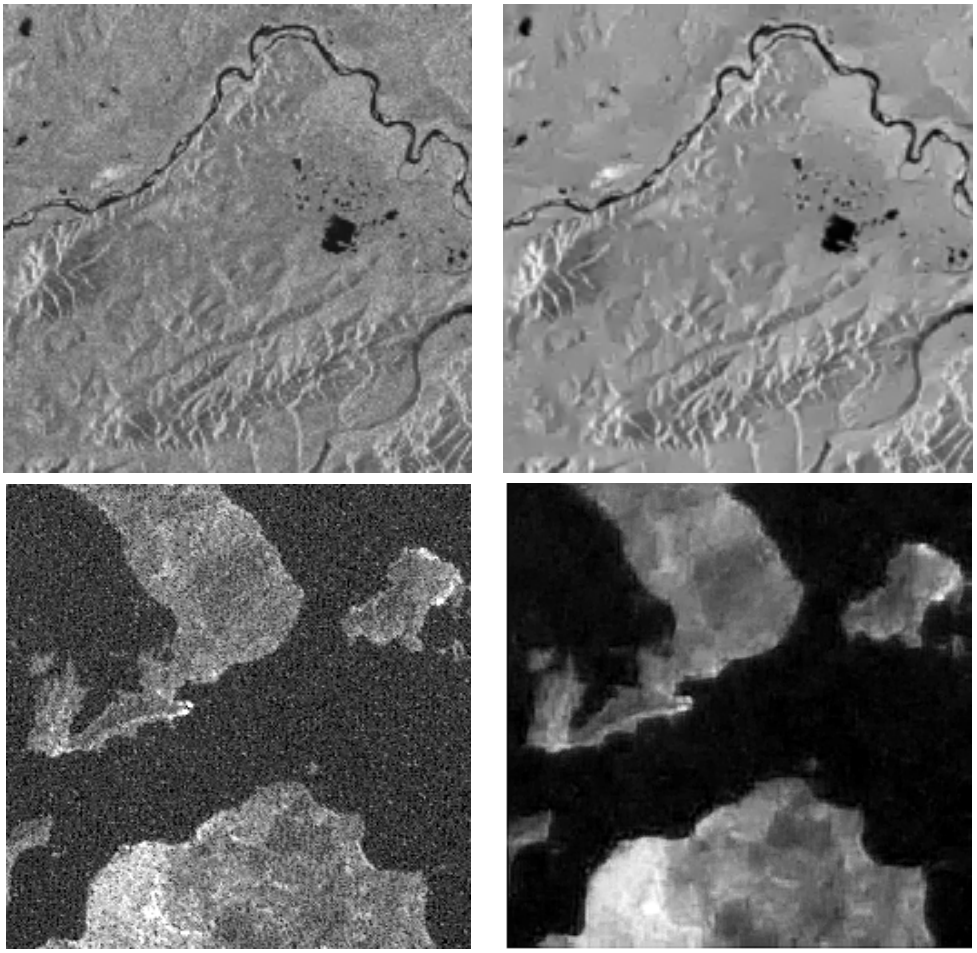


FIGURE 6.3. *Left: original SAR images. Right: the corresponding results of the Bayesian wavelet shrinkage using the MAP-MRF approach.*

Finally, we apply the same approach to a third kind of images, i.e., medical ultrasound images. There, noise is again of the speckle type, but the speckle is different than in the previous case. Ultrasound images are also special in the sense that it is difficult to judge the quality of the results. Only medical experts are competent to decide which image details are relevant, and which regions can be smoothed without loosing important information. Therefore, it is very desirable for the denoising approach in this case to be user-interactive, so that the amount of smoothing in "flat" areas can be easily changed and controlled; precisely this is what the Bayesian shrinkage using MRF priors offers (see Fig. 5.2 and the related explanation). In Fig. 6.4 we show an original ultrasound image and two results of the same method, but with increasing influence of the prior model from left to right. One can see that white "clouds" of speckles become smoother as the influence of the prior model increases.

7. Conclusions. In this chapter, we have discussed a Bayesian approach to wavelet based image denoising using MRF prior models. Two different directions in this respect were analyzed: the classical MAP-MRF estimation applied in the wavelet

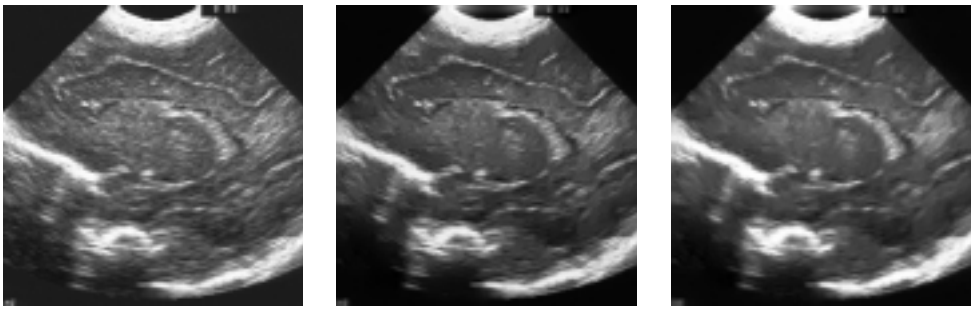


FIGURE 6.4. *Left: an original ultrasound image. Middle and Right: two results of the Bayesian wavelet shrinkage using the MAP-MRF approach, with increasing influence of the prior.*

domain, and the use of binary MRF priors in wavelet shrinkage techniques. The first approach theoretically offers better results, while the second one leads to much faster algorithms and seems to be attractive for many practical applications.

For the first approach, practical results were not available yet. However, some practical aspects regarding the computation, possible advantages as well as the specific problems were discussed. This approach is related to well studied single-resolution [32] and multigrid [33] MAP-MRF estimation methods. On the other hand, its implementation in the wavelet domain opens new possibilities. In particular, new types of clique potentials, which better adapt to the actual image edges should be considered. For detecting the meaningful edges in the presence of noise, the wavelet representation is especially convenient. In this framework, the construction of different edge adaptive potentials is also facilitated.

The second approach applies the MAP-MRF estimation of masks that indicate the positions of significant wavelet coefficients, and accordingly adapts the wavelet shrinkage to the actual image edges. In contrast to standard techniques, which apply the same shrinkage nonlinearity to all the coefficients in a detail image, this approach offers a family of shrinkage functions, depending on the spatial surrounding. The practical importance is the following: one can preserve the connected edge clusters of the wavelet coefficients, and at the same suppress even larger, but isolated ones. The results demonstrate that in this case the important image details are better reconstructed from noise. Furthermore, the results show that this approach is very efficient for suppressing different kinds of natural noise, like in infrared, SAR and ultrasound images. An important advantage is that it allows a user to easily change the amount of smoothing in flat areas without edges. It is simply done by changing the relative importance given to the measured data at one hand, and to the prior knowledge about spatial clustering at the other. This is, for example, very interesting for denoising of medical ultrasound images, where clouds of speckles may contain useful information, and only experienced users may decide to which amount those areas should or may be smoothed.

The obtained results motivate further research in this field. For example, the improvements can be expected from defining specific clique potential functions for different detail images. These potentials should give preference to relatively narrow clusters of edge labels in predominantly horizontal, vertical or diagonal direction, depending on the type of detail image. In addition, for different kinds of noise one should examine the optimal choice of the significance measure (e.g., the magnitude of the wavelet coefficient, the interscale ratio, the interscale product, etc). For the chosen

measure the realistic likelihood models, given noise and given useful edges should be derived. We have presented in [58] some current developments in this respect.

Acknowledgment. This research was performed in the scope of the Belgian humanitarian demining project, HUDEM. The authors thank to other partners in this project for providing the row data. The authors also thank to the anonymous reviewer, for many useful and constructive suggestions. This work was partially supported by the Fund for Scientific Research - Flanders, Belgium (FWO - project G.0037.00).

REFERENCES

[1] K. Abend, T. J. Harley, and L. N. Kanal, *"Classification of binary random patterns,"* IEEE, Trans. Inform. Theory, IT-11 (1965), pp. 538–544.
[2] N. Ahuja, A. Rosenfeld, *"Mosaic Models for Textures,"* IEEE, Trans. Pattern Anal. and Machine Intel., 3 (1981), pp. 1–11.
[3] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, *"Image Coding Using Wavelet Transform,"* IEEE, Trans. Image Proc., 1 (1992), pp. 205–220.
[4] M. R. Banham, A. K. Katsaggelos, *"Spatially Adaptive Wavelet-Based Multiscale Image Restoration,"* Trans. Image Proc. 5, (1996), pp. 619-634.
[5] J. Besag, *"Spatial Interaction and Statistical Analysis of Lattice Systems,"* J. Royal Statistical Society, Ser. B, 6 (1974), pp. 192–236.
[6] J. Besag, *"On the Statistical Analysis of Dirty Pictures,"* J. Royal Statistical Society, Ser. B, 48 (1986), pp. 259–302.
[7] A. Blake, *"Comparisson of the Efficiency of Deterministic and Stochastic Algorithms for Visual Reconstruction,"* IEEE, Trans. Pattern Anal. and Machine Intel., 11 (1989), pp. 2–12.
[8] C. Bouman and B. Liu, *"Multiple Resolution Segmentation of Textured Images,"* IEEE, Trans. Pattern Anal. and Machine Intel., 13 (1991), pp. 99–113.
[9] R. Birk, W. Camus, E. Valenti, and W. McCandless, *"Synthetic aperture radar imaging systems,"* IEEE AES Systems Magazine, 10 (1995), pp. 15–23.
[10] J. Canny, *"A computational Approach to Edge Detection,"* IEEE, Trans. Pattern Anal. and Machine Intel., 8 (1986), pp. 679–698.
[11] A. E. Cetin and R. Ansari, *"Signal Recovery from Wavelet Transform Maxima,"* IEEE, Trans. Signal Proc. 42 (1994), pp. 194–196.
[12] A. Chambolle, R. A. DeVore, N.-y. Lee, and B. J. Lucier, *"Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal Through Wavelet Shrinkage,"* IEEE, Trans. Image Proc. 7 (1998), pp. 319–335.
[13] R. Chellappa, Y.-H. Hu, and S.-Y. Kung, *"On two-dimensional Markov Spectral Estimation,"* IEEE, Trans. Acoustics, Speech, Signal Proc., 31 (1983), pp. 836–841.
[14] R. Chellappa, *"Two-dimensional Discrete Gaussian Markov Random Field Models for Image Processing,"*, Progress in Pattern Recognition, Edited by L. N. Kanal and A. Rosenfeld, North-Holland 1985.
[15] T.-C. Hsung, D. P.-K. Lun, and W. -C. Siu, *"Denoising by Singularity Detection,"* IEEE, Trans. Signal Proc. 47 (1999), pp. 3139–3144.
[16] A. Cohen and J. Kovacevic, *"Wavelets: The Mathematical Background,"*, Proceedings of the IEEE, 84 (1996), pp. 514–522.
[17] M. L. Comer and E. J. Delp, *"Segmentation of Textured Images Using a Multiresolution Gaussian Autoregressive Model,"* IEEE, Trans. Image Proc., 8 (1999), pp. 408–420.
[18] M. S. Crouse, R. D. Nowak, and R. G. Baranuik, *"Wavelet-Based Statistical Signal Processing Using Hidden Markov Models,"* IEEE Trans. Signal Proc., 46 (1998), pp. 886–902.
[19] G. R. Cross and A. K. Jain, *"Markov Random Field Texture Models,"* IEEE, Trans. Pattern Anal. and Machine Intel., 5 (1983), pp. 25–39.
[20] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM (1992).
[21] H. Derin, H. Elliott, R. Cristi, and D. Geman, *"Bayes Smoothing Algorithms for Segmentation of Binary Images Modeled by Markov Random Fields,"* IEEE, Trans. Pattern Anal. and Machine Intel., 6 (1984), pp. 707–720.
[22] H. Derin and H. Elliott, *"Modeling and Segmentation of Noisy and Texture Images Using Gibbs Random Fields,"* IEEE, Trans. Pattern Anal. and Machine Intel., 9 (1987), pp. 39–55.

[23] D. L. Donoho and I. M. Johnstone, *"Ideal spatial adaptation by wavelet shrinkage,"* Technical Report, Statistics, Stanford (1992).
[24] D. L. Donoho and I. M. Johnstone, *"Minimax Estimation via Wavelet Shrinkage,"* Technical Report, Statistics, Stanford (1992).
[25] D. L. Donoho, *"De-Noising by Soft-Thresholding,"* IEEE, Trans. Inform. Theory, 41 (1995), pp. 613–627.
[26] D. L. Donoho and I. M. Johnstone, *"Adapting to unknown smoothness via wavelet shrinkage,"* J. Amer. Stat. Assoc., 90, (1995).
[27] R. C. Dubes and A. K. Jain, *"Random Field Models in Image Analysis,"* Progress in Pattern Recognition, Edited by L. N. Kanal and A. Rosenfeld, North-Holland 1985.
[28] G. Fan and X.-G. Xia, *"Image Denoising Using Local Contextual Hidden Markov Model in the Wavelet Domain,"* submitted to IEEE, Signal Proc. Letters., (2000).
[29] G. Fan and X.-G. Xia, *"Wavelet-Based Image Denoising Using Hidden Markov Models,"* Proc. of the IEEE International Conf. on Image Proc. (ICIP), Vancouver, BC, Canada, (2000).
[30] Y. Fisher (editor), *Fractal Image Compression: Theory and Application*, Springer Verlag, (1995), New York.
[31] M. A. T. Figueiredo, J. M. N. Leitão , *"Unsupervized Image Restoration and Edge Location Using Compound Gauss-Markov Random Fields and the MDL Principle,"* IEEE, Trans. Image Proc. 6 (1997), pp. 1089–1102.
[32] S. Geman, D. Geman, *"Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images,"* IEEE, Trans. Pattern Anal. and Machine Intel., 6 (1984) pp. 721–741.
[33] B. Gidas, *"A Renormalization Group Approach to Image Processing Problems,"*, IEEE, Trans. Pattern Anal. and Machine Intel., 11 (1989), pp. 164–180.
[34] J. W. Goodman, "Some Fundamental Proprerties of Speckle," *J. Opt. Soc. Am.*, vol. 66, pp. 1145–1150, 1976.
[35] M. L. Hilton and R. T. Ogden, *"Data Analytic Wavelet Threshold Selection in 2-D Denoising,"* IEEE, Trans. Signal Proc., 45 (1997) pp. 496–500.
[36] T.-C. Hsung, D. P.-K Lun, and W.-C. Siu, *"Denoising by singularity detection,"* IEEE Trans. Signal Proc., 47 (1999), pp. 3139–3144.
[37] A. K. Jain, *"Advances in Mathematical Models for Image Processing,"*, Proceedings IEEE, 69 (1981) pp. 502–528.
[38] M. Jansen, M. Malfait, A. Bultheel, *"Generalized Cross Validation for Wavelet Thresholding,"*, Signal Processing, 56 (1997) pp. 33–44.
[39] R. L. Kashiap and R. Chellappa, *"Estimation and Choice of Neighbors in Spatial-Interaction Models of Images,"* IEEE, Trans. Inform. Theory, 29 (1983), pp. 60–72.
[40] Z. Kato, M. Berthod, and J. Zerubia, *"A Hierarchical Markov Random Field Model and Multi-Temperature Annealing for Parallel Image Classification ,"* Rapport de recherche No 1938, Institut National de Recherche en Informatique, Robotique, image et vision (1994).
[41] A. Khotanzad and J. Bennett, *"A Spatial Correlation Based Method for Neighbor Set Selection in Random Field Image Models,"* IEEE, Trans. Image Proc., 8 (1999), pp. 734–740.
[42] M. Kivanc Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, *"Low-Complexity Image Denoising Based on Statistical Modeling of Wavelet Coefficients,"* IEEE, Signal Proc. Letters, 6 (1999), pp. 300–303.
[43] S. Krishnamachari and R. Chelappa, *"Multiresolution Gauss-Markov Random Field Models for Texture Segmentation,"* IEEE, Trans. Image Proc, 6 (1997), pp. 251–267.
[44] S. Z. Li, *Markov Random Field Modeling in Computer Vision,* Springer-Verlag (1995)
[45] M. R. Luettgen, W. C. Karl, A. S. Willsky, and R. R. Tenney, *"Multiscale Representations of Markov Random Fields,"* IEEE, Trans. Signal Proc., 41 (1993), pp. 3377–3396.
[46] S. Mallat, *"Multifrequency Channel Decomposition of mages and Wavelet Models,"* IEEE, Trans. Acoustics, Speech, Signal Proc., 37 (1989), pp. 2091–2110.
[47] S. Mallat, *"A theory for Multiresolution Signal Decomposition: The Wavelet Representation,"* IEEE, Trans. Pattern Anal. and Machine Intel., 11 (1989), pp. 674–693.
[48] S. Mallat, W. L. Hwang, *"Singularity Detection and Processing with Wavelets,"* IEEE Trans. Inform. Theory, 38 (1992) pp. 617–643.
[49] S. Mallat, S. Zhong, *"Characterization of Signals from Multiscale Edges,"* IEEE, Trans. Pattern Anal. and Machine Intel, 14 (1992), pp. 710–732.
[50] S. Mallat, *A wavelet tour of signal processing*, Academic Press (1998).
[51] M. Malfait, D. Roose, *"Wavelet-Based Image Denoising Using a Markov Random Field a Priori Model,"* IEEE, Trans. Image Proc., 6, (1997) pp. 549–565.
[52] V. A. Malyshev and R. A. Minlos, *Gibbs Random Fields. Method of cluster expansions*, Kluwer, Academic Publishers (1991).

[53] P. Moulin and J. Liu, *"Analysis of Multiresolution Image Denoising Schemes Using Generalized Gaussian and Complexity Priors,"* IEEE, Trans. Inform. Theory, 45 (1999), pp. 909–919.

[54] G. P. Nason, *"Wavelet Regression by Cross-Validation,"* Technical Report 447, Department of Statistics, Stanford (1994).

[55] A. P. Pentland, *"Fractal-Based Description of Natural Scenes,"* IEEE, Trans. Pattern Anal. and Machine Intel., 6 (1984), pp. 661–674.

[56] A. Pizurica, W. Philips, I. Lemahieu, M. Acheroy, *"Image De-Noising In The Wavelet Domain Using Prior Spatial Constraints,"* Proc. of the IEE Conf. on Image Processing and its Applications (IPA), Manchester, UK, (1999), pp. 216–219.

[57] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, *"A Wavelet-Based Image Denoising Technique Using Spatial Priors,"* Proc. of the IEEE International Conf. on Image Proc. (ICIP), Vancouver, BC, Canada, (2000), pp. 296–299.

[58] A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy, *"Image Denoising via Wavelet Shrinkage Using a Prior Model for Spatial Clustering"*, Technical Report, Dept. of Telecommunications and Information Processing, Ghent University, Ghent, Belgium. Submitted to IEEE Trans. Image Proc.

[59] J. Romberg, H. Choi, R. Baraniuk, and N. Kingsbury, *Multiscale Classification Using Complex Wavelets and Hidden Markov Tree Models ,"* Proc. of the IEEE International Conf. on Image Proc. (ICIP), Vancouver, BC, Canada, (2000).

[60] J. M. Shapiro, *"Embedded Image Coding Using Zerotrees of Wavelet Coefficients,"* IEEE, Trans. Signal Proc., 41 (1993), pp. 3445–3462.

[61] E. P. Simoncelli and E. H. Adelson, *"Noise Removal via Bayesian Wavelet Coring,"* in Proc. IEEE International Conf. Image Processing (ICIP), Lausanne, Switzerland (1996), pp. 379–382.

[62] E. P. Simoncelli, *"Modeling the joint statistics of image in the wavelet domain,"* in Proc. SPIE Conf. 3813 on Wavelet Applications in Signal and Image Processing VII, Denver, CO, (1999).

[63] P. C. Smits, S. G. Dellepiane, and G. Vernazza, *"Discontinuity Adaptive MRF Model for Synthetic Aperture Radar Image Analysis,"* In Proc. International Conf. on Image Analysis and Processing (ICIAP) I (1997), 255–262

[64] C. Taswell, *"The What, How and Why of Wavelet Shrinkage Denoising,"* Computational Toolsmiths, Stanford, CA 94309-9925, Technical Report CT-1998-09 (1998).

[65] B. Vidakovic, *"Nonlinear Wavelet Shrinkage with Bayes Rules and Bayes Factors,"* Duke Univ., Durham, NC, (1994), preprint.

[66] Y. Xu, J. B. Weaver, D. M. Healy, J. Lu, *"Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique,"* IEEE Trans. Image Proc. 3 (1994) pp. 747–758.

[67] N. Weyrich and G. T. Warhola, *"Wavelet Shrinkage and Generalized Cross Validation for Image Denoising,"* IEEE, Trans. Image Proc., 7 (1998), pp. 82–90.

[68] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Caro Methods* , Springer-Verlag (1995).