

Stochastic Nonlinear Image Restoration Using the Wavelet Transform

Marc C. Robini, *Member, IEEE*, and Isabelle E. Magnin, *Member, IEEE*

Abstract—The dominant methodology for image restoration is to stabilize the problem by including a roughness penalty in addition to faithfulness to the data. Among various choices, concave stabilizers stand out for their boundary detection capabilities, but the resulting cost function to be minimized is generally multimodal. Although simulated annealing is theoretically optimal to take up this challenge, standard stochastic algorithms suffer from two drawbacks: i) practical convergence difficulties are encountered with second-order prior models and ii) it remains computationally demanding to favor the formation of smooth contour lines by taking the discontinuity field explicitly into account. This work shows that both weaknesses can be overcome in a multiresolution framework by means of the 2-D discrete wavelet transform (DWT). We first propose to improve convergence toward global minima by single-site updating on the wavelet domain. For this purpose, a new restricted DWT space is introduced and a theoretically sound updating mechanism is constructed on this subspace. Next, we suggest to incorporate the smoothness of the discontinuity field via an additional penalty term defined on the high frequency subbands. The resulting increase in complexity is small and the approach requires the specification of a unique extra parameter for which an explicit selection formula is derived.

Index Terms—Image restoration, inverse problems, metropolis dynamics, simulated annealing, wavelets.

NOMENCLATURE

BSNR	Blurred signal-to-noise ratio.
DWT	Discrete wavelet transform.
ISNR	Improvement in signal-to-noise ratio.
LB	“Locally bounded.”
WaRD	Wavelet-based regularized deconvolution.
α	Concavity parameter of the ϕ -function.
(β_n)	Cooling schedule.
Δ	Scale parameter.
η	Additive, zero-mean, Gaussian white noise field corrupting the blurred image.
$\tilde{\lambda}$	Image smoothing parameter.
$\tilde{\lambda}$	Wavelet smoothing parameter.
$\lambda^*, \tilde{\lambda}^*$	Lower bounds for λ and $\tilde{\lambda}$, respectively.
σ	Standard deviation of η .
ϕ	“Potential function.”
$\Phi^{(k)}$	k th-order prior.

$\tilde{\Phi}$	Wavelet smoothing term.
$\Omega^{l S }$	LB image subspace.
$\tilde{\Omega}_J^{l S }$	Restricted wavelet space.
$\mathcal{A}(\cdot)$	Annealing on $\Omega^{l S }$.
$\tilde{\mathcal{A}}(\cdot)$	Annealing on $\tilde{\Omega}_J^{l S }$.
\mathfrak{A}_J	DWT analysis matrix.
d	Degraded observation (data).
$\mathcal{D}^{(k)}$	k th-order discrete derivative operator.
\mathcal{D}_j^i	Gradient operator in a vertical ($i = 1$) or horizontal ($i = 2$) direction.
\mathcal{H}	Blurring operator.
J	Wavelet decomposition level.
$\mathcal{N}, \mathcal{N}^i$	Neighborhood structures (a neighborhood system \mathcal{N} on a set \mathcal{C} is denoted by $\mathcal{N}(\mathcal{C})$ and $\mathcal{N}_e(\mathcal{C})$ stands for the set of neighbors of $e \in \mathcal{C}$).
q', \tilde{q}'	Communication kernels on $\Omega^{l S }$ and $\tilde{\Omega}_J^{l S }$, respectively.
r_J	Low resolution residual.
S, S_J, S_J^i	Rectangular lattices respectively supporting x^0 , r_J and w_j^i .
\mathfrak{S}_J	DWT synthesis matrix.
$U^{(k)}$	k th-order posterior cost function defined on the image space.
$\tilde{U}_J^{(2)}$	Second-order posterior cost function defined on the wavelet domain.
V	Cost function $U^{(1)} + \tilde{\lambda}\tilde{\Phi}$.
w_j^i	Detail subband with orientation i at decomposition level j .
x^0	Original intensity distribution.
$x_{\setminus s}$	Configuration x with x_s removed.

I. INTRODUCTION

THE discrete image restoration problem is to recover an original intensity distribution $x^0 = \{x_s^0; s \in S\}$ defined over a 2-D rectangular lattice S from its measurement $d = \{d_s; s \in S' \subset S\}$ degraded by the sensing environment. In many situations, the transformation from x^0 to d is well described by the familiar additive linear degradation model $d_s = (\mathcal{H}x^0)_s + \eta_s$, where \mathcal{H} is a linear operator representing a (possibly space-variant) point spread function and the noise process $\eta = \{\eta_s; s \in S'\}$ consists of independent and identically distributed zero-mean Gaussian random variables. In the following, we assume that both \mathcal{H} and the noise variance σ^2 are known.

Using lexicographic ordering, we obtain the underdetermined linear system $d = \mathcal{H}x^0 + \eta$, where $x^0 \in \mathbb{R}^{|S| \times 1}$, $d, \eta \in \mathbb{R}^{|S'| \times 1}$ and $\mathcal{H} \in \mathbb{R}^{|S'| \times |S|}$. Owing to the ill-posedness of the initial continuous problem (a particular case of a first-kind Fredholm in-

Manuscript received January 8, 2002; revised January 29, 2003. This material was presented in part at the IEEE International Conference on Image Processing, Chicago, IL, in October 1998. This work has been performed in accordance with the scientific trends of the national research group GDR ISIS of the CNRS. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mario A. T. Figueiredo.

The authors are with CREATIS, INSA de Lyon, 69621 Villeurbanne cedex, France (e-mail: marc.robini@creatis.insa-lyon.fr).

Digital Object Identifier 10.1109/TIP.2003.812330

tegral equation), the normal matrix $\mathcal{H}^T \mathcal{H}$ is poorly conditioned and hence the error propagation level from the data d to the solution is most often unacceptable. In order to circumvent this difficulty, many image restoration methods include prior constraints about the true image x^0 in addition to those implicit in coherence to the data. A common estimate of x^0 that arises from bayesian modeling [1], [2] or, equivalently, from a likelihood model with regularization [3], [4], is defined as any global minimum of a cost function

$$U^{(k)} : \mathbb{R}^{|S|} \rightarrow \mathbb{R} \\ x \mapsto \|\mathcal{H}x - d\|_2^2 + \lambda \Phi^{(k)}(x) \quad (1)$$

where the stabilizer $\Phi^{(k)}$ promotes the formation of smooth regions separated by $(k - 1)$ th-order discontinuities and $\lambda \in \mathbb{R}_+^*$ governs the trade-off between the degree of stabilization of the solution and its faithfulness to the data. Confining ourselves to approximately piecewise constant ($k = 1$) or planar ($k = 2$) original distributions, we consider the widespread model

$$\Phi^{(k)}(x) = \sum_{l=1}^{L_k} \phi \left(\frac{[\mathcal{D}^{(k)}x]_l}{\Delta} \right) \quad (2)$$

where $\phi : \mathbb{R} \rightarrow \mathbb{R}^+$ is taken to be even, increasing in \mathbb{R}_+ and such that $\phi(0) = 0$, $\mathcal{D}^{(k)} \in \mathbb{R}^{L_k \times |S|}$ is the matrix implementation of a k th-order discrete derivative operator, and $\Delta \in \mathbb{R}_+^*$ is a scale parameter. For the first-order case, L_1 is the number of two-sites cliques $\{s, t\} \subset S$ associated with the eight-nearest neighbor system. Let l_c be the row index of $\mathcal{D}^{(1)}$ corresponding to a given clique c . Then $[\mathcal{D}^{(1)}x]_{l_c} = x_s - x_t$ if c is horizontal or vertical and $[\mathcal{D}^{(1)}x]_{l_c} = (x_s - x_t)/\sqrt{2}$ if c is diagonal. For the second-order case, the summation is over three- and four-sites cliques of the form

$$\begin{array}{ccc} \text{(i)} \begin{array}{ccc} \bullet^s & & \bullet^t \\ \bullet^s & & \bullet^t \\ \bullet^u & & \bullet^v \end{array} & \text{(ii)} \begin{array}{ccc} \bullet^s & & \bullet^t \\ \bullet^s & & \bullet^t \\ \bullet^u & & \bullet^v \end{array} & \text{(iii)} \begin{array}{ccc} \bullet^s & & \bullet^t \\ \bullet^s & & \bullet^t \\ \bullet^u & & \bullet^v \end{array} \end{array}$$

and $\mathcal{D}^{(2)}$ is defined by $[\mathcal{D}^{(2)}x]_{l_c} = x_s - 2x_t + x_u$ if c is of type (i) or (ii) and $[\mathcal{D}^{(2)}x]_{l_c} = x_s - x_t - x_u + x_v$ if c is of type (iii).

The form of the ‘‘potential function’’ ϕ in (2) plays a crucial role in the estimation process. In recent years, deriving conditions for the ϕ -function to preserve or to detect discontinuities has been an active research area [5]–[10]. On one hand, some authors [6], [7], [11]–[15] encourage the use of convex ϕ -functions to ensure the convexity of $U^{(k)}$ while preserving edges in the sense that smoothing is reduced in their vicinity. On the other hand, one can be interested in nonconvex ϕ -functions [5], [16]–[19] which yield sharper edges and allow to achieve better resolution [8] at the expense of instability. In accordance with the latter standpoint, following Geman and Reynolds [5], we use functions of the form

$$\phi(w) = \frac{|w|}{(1 + |w|)^\alpha}, \quad \alpha \in (0, 1] \quad (3)$$

which are strictly concave in \mathbb{R}_+ . For any $\alpha \in (0, 1]$, one very important property of ϕ is that it admits a strictly positive right derivative at zero ($\lim_{w \rightarrow 0^+} w^{-1} \phi(w) = 1$). It follows that almost any strict minimizer of $U^{(k)}$ not only recovers regions

where the k th-order differences are null but is also stable with respect to small variations of the data [20].

Because of the concavity of ϕ , $U^{(k)}$ is most often nonconvex and typically exhibits numerous local minima. The minimization of $U^{(k)}$ is thus nontrivial and one can employ either sub-optimal algorithms such as iterated conditional modes [2] and graduated nonconvexity [19], [21], [22] or stochastic relaxation with annealing which has the remarkable property to be asymptotically optimal when properly tuned [23]–[25]. The stochastic approach, which is considered in this paper, gives satisfactory results using finite-time exponential cooling schedules and carrying out the minimization on a locally bounded image subspace [26]. However, such algorithms still suffer from two drawbacks:

- 1) Markov chain Monte Carlo algorithms defined on a large state space with loosely coupled sites experience difficulties when pixels are updated singly (see, e.g., Jennison’s discussion in [27]). More specifically, the convergence is slow when the low cost regions of the state space correspond to configurations having a strong low frequency content. This is particularly true for distributions involving a second-order prior such as $U^{(2)}$, a common practical situation since many images are better described by planar areas separated by roof-edges than by constant regions [5], [28]. Indeed, second-order models produce more complex energy landscapes than first-order models, which translates to a less stable inversion process.
- 2) In the case of a first-order model, the minimization of $U^{(1)}$ leads to noisy object boundaries which are not faithful to the original distribution. This behavior is quite predictable as the model does not embed any prior knowledge on the geometry of edges. Unfortunately, the mutual dependence between neighboring discontinuities is difficult to take into account unless an explicit line process is introduced [1], [3], [29], which results in a severe increase in computational complexity. Furthermore, the penalties for the various edge configurations to be specified then are so many additional hyper-parameters to deal with.

Our concern is to show that the above difficulties can be alleviated in a multiresolution framework by means of the 2-D discrete wavelet transform (DWT):

- 1) Since single-site updating on the spatial frequency space corresponds to some block-site updating on the spatial domain, wider site connections are achieved by minimizing the composite of the inverse DWT and $U^{(2)}$. This approach is conceptually simple and fundamentally different from the renormalization group method [30] as well as multi-grid Monte Carlo algorithms [31]; it is based on the obvious fact that the set of true image estimates $\{x \in \mathbb{R}^{|S|} \mid U^{(2)}(x) = \min_z U^{(2)}(z)\}$ also writes $\{x \in \mathbb{R}^{|S|} \mid U^{(2)}(x) = \min_{\tilde{x}} U^{(2)}(\mathfrak{S}\tilde{x})\}$, where \mathfrak{S} is the synthesis matrix associated with the transform under consideration. In order to achieve efficient sampling of the state space, we introduce a restricted DWT space, say $\tilde{\Omega}$, with finite discrete coefficients and locally bounded subbands. In particular, we show that it is possible to construct a single-site updating dynamics on $\tilde{\Omega}$ that preserves

the theoretical convergence properties of annealing. In comparison with annealing on the digital image space, the proposed inversion process is more stable and our experiments show significant benefits in terms of both final energy level and improvement in signal-to-noise ratio.

- 2) In order to encourage the formation of smooth contours, we suggest to incorporate the smoothness features of the line field implicitly rather than explicitly. To be more precise, we consider the minimization of an augmented functional

$$V : \mathbb{R}^{|S|} \rightarrow \mathbb{R}$$

$$x \mapsto U^{(1)}(x) + \tilde{\lambda} \Psi(\mathcal{A}x)$$

where \mathcal{A} is a DWT analysis matrix, Ψ is a penalty term operating on horizontal and vertical detail subbands by means of gradient operators, and $\tilde{\lambda} \in \mathbb{R}_+^*$ weights the influence of Ψ . To avoid the introduction of an additional free parameter, we use some ideas introduced in [5] to prove a lower bound on $\tilde{\lambda}$ for the true image x^0 to be a coordinate-wise minimum of V with probability greater than some specified threshold value. This bound is expressed as a function of λ, Δ, α , the sum of squares of the blur coefficients (assuming \mathcal{H} is space-invariant) and the noise standard deviation. Our restoration results show that the proposed method yield state-of-the-art performance under the considered experimental conditions.

The paper is organized as follows. Metropolis-type annealing algorithms and their “conventional” application to image restoration are briefly discussed in the next section. Section III is devoted to image restoration by annealing on a restricted wavelet space. In Section IV, we describe our solution to the issue of edge continuation along with the choice of $\tilde{\lambda}$. Experimental results appear in Section V and conclusions are presented in Section VI.

II. BACKGROUND

A. Annealing in a Few Words

Let us consider an energy landscape (E, U, q) , where E is a finite set called the configuration space or state space, $U : E \rightarrow \mathbb{R}$ is a real-valued function to be minimized on E , and $q : E \times E \rightarrow [0, 1]$ is a symmetric and irreducible Markov matrix (i.e., $\sum_{z \in E} q(x, z) = 1$, $q(x, y) = q(y, x)$ and $\sup_{n \in \mathbb{N}^*} q^n(x, y) > 0$ for all $x, y \in E$). In most practical situations, the “communication kernel” q is defined by selecting a neighborhood system $\mathcal{N}(E) = \{\mathcal{N}_x(E); x \in E\}$ and setting

$$q(x, y) = \mathbf{1}_{\{y \in \mathcal{N}_x(E)\}} |\mathcal{N}_x(E)|^{-1}. \quad (4)$$

For any $\beta \in \mathbb{R}_+^*$, define the transition probability matrix P_β on E by

$$P_\beta(x, y) = \begin{cases} q(x, y) \exp\left(-\beta(U(y) - U(x))^+\right) & \text{if } y \neq x, \\ 1 - \sum_{z \neq x} P_\beta(x, z) & \text{if } y = x \end{cases} \quad (5)$$

where $a^+ := \max\{a, 0\}$, and let $(\beta_n)_{n \in \mathbb{N}^*}$ be a nondecreasing positive real sequence called the cooling schedule. A Metropolis-type annealing algorithm on (E, U, q) is a discrete-time nonhomogeneous Markov chain $(X_n)_{n \in \mathbb{N}}$ with transitions $P(X_n = y | X_{n-1} = x) = P_{\beta_n}(x, y)$, $x, y \in E$. We shall use the notation $\mathbf{M}(E, U, q, (\beta_n))$ for short. Under our assumptions, P_β is irreducible and aperiodic. Its unique equilibrium probability measure is the Gibbs distribution π_β with energy U at temperature β^{-1} defined by $\pi_\beta(x) = Z^{-1} \exp(-\beta U(x))$, $x \in E$, where Z is a normalizing constant. It is easy to check that π_β tends to the uniform distribution on the set E_{\min} of global minima of U as β tends to infinity. Hence, the key idea of annealing is that, for sufficiently slowly increasing cooling schedule, the law of X_n should be close to π_{β_n} and, consequently, one can expect that

$$\lim_{n \rightarrow \infty} \inf_{x \in E} P(X_n \in E_{\min} | X_0 = x) = 1.$$

Early results [1], [23]–[25] show that this desirable property holds for suitably adjusted logarithmic schedules. Still, it has been rigorously demonstrated [32] and experimentally verified [26] that the associated algorithms generally perform poorly as soon as one deals with a finite number of iterations and that exponential schedules should be preferred.

B. Conventional Annealing-Based Image Restoration

In straightforward applications of Metropolis-type annealing to image restoration, the state space is of the form $\Lambda^{|S|}$, where $\Lambda = \{al + b; l = 0, \dots, \mathcal{Q} - 1\}$, $\mathcal{Q} \in \mathbb{N} \setminus \{0, 1\}$, $a \in \mathbb{R}_+^*$, $b \in \mathbb{R}$, and a common choice for the communication kernel (4) is defined by

$$\mathcal{N}_x(\Lambda^{|S|}) = \left\{ y \in \Lambda^{|S|} \mid (\exists! s \in S) [y_s \neq x_s] \right\}.$$

However, because the proposed new pixel values are uniformly generated over Λ , $P(X_n \neq X_{n-1})$ quickly gets close to zero as β_n increases. It follows that outlier intensity values frequently subsist and that image boundaries are difficult to alter. As shown in [33] and [26], this problem can be overcome by restricting $\Lambda^{|S|}$ to a locally bounded (LB) image subspace, that is, a set of configurations $x \in \Lambda^{|S|}$ such that, for all $s \in S$

$$\min_{t \in \mathcal{N}_s(S)} x_t - \zeta a \leq x_s \leq \max_{t \in \mathcal{N}_s(S)} x_t + \zeta a \quad (6)$$

where $\zeta \in \{1, \dots, \mathcal{Q} - 1\}$ and $\mathcal{N}(S)$ is a predefined neighborhood system on S (a typical choice is a four-nearest neighbor system together with $\zeta \ll \mathcal{Q}$, e.g., $\zeta = 5$ if $\mathcal{Q} = 256$). Such a set is denoted by $\Omega(\zeta, \Lambda, \mathcal{N}(S)) =: \Omega^{|S|}$ and the corresponding single-site updating Metropolis dynamics can be briefly described as follows. Given $x \in \Omega^{|S|}$, let $x_{\setminus s}$ be the configuration x with x_s removed and let $x_{\setminus s}(\mathcal{L})$ be the configuration defined by $(x_{\setminus s}(\mathcal{L}))_t = \mathbf{1}_{\{t=s\}} \mathcal{L} + \mathbf{1}_{\{t \neq s\}} x_t$. We denote the section of $\Omega^{|S|}$ at $x_{\setminus s}$ by $\gamma_s(x)$, that is, $\gamma_s(x) = \{\mathcal{L} \in \Lambda \mid x_{\setminus s}(\mathcal{L}) \in \Omega^{|S|}\}$. Then, for all $x, y \in \Omega^{|S|}$ such that $x \neq y$, the communication kernel q' on $\Omega^{|S|}$ is defined by

$$q'(x, y; \zeta, \Lambda, \mathcal{N}(S)) = \begin{cases} (|S| \cdot |\gamma_s(x)|)^{-1} & \text{if } y \in \Gamma(x, s), \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where $\Gamma(x, s)$ is the set of configurations z satisfying $z_{\setminus s} = x_{\setminus s}$ and $z_s \in \gamma_s(x) \setminus \{x_s\}$. Reconsidering (6), one can find the answer to what makes this dynamics appropriate to piecewise smooth image recovery. At the beginning of the annealing process, that is, when β_n is low, the visited states are noisy and thus γ_s is wide for all $s \in S$. Conversely, as β_n increases, a smoother image is gradually formed so that γ_s becomes smaller at most sites. In other words, adequate sampling of the state space is allowed at high temperatures while the generation of very unlikely candidates is avoided at low temperatures. Another strong point is that q' can be shown to be symmetric and irreducible [33]. Therefore, the convergence results mentioned in Section II-A hold for $\mathbf{M}(\Omega^{|S|}, U, q', (\beta_n))$.

III. IMAGE RESTORATION IN A WAVELET BASIS

Depending on the cost function U to be minimized, the annealing algorithms $\mathbf{M}(\Omega^{|S|}, U, q', (\beta_n))$ can still show mediocre performance despite the benefits resulting from the restriction to an LB subspace. A case in point is $U = U^{(k)}$ (1)–(2) for $k = 2$, as second-order models produce “perverse” energy landscapes whose structure strongly limits the moves between high probability areas of the state space separated by low probability regions. An immediate solution to this problem is to generate new candidates by changing several pixel values simultaneously, which can be implicitly carried out by means of single-site updating on the wavelet domain.

In a wavelet basis, an image x is represented by a set of subimages $\{r_J, w_j^i; j = 1, \dots, J, i = 1, 2, 3\}$, where r_J is the approximation of x at resolution 2^{-J} and the spatially oriented high-frequency (HF) subbands $\{w_j^i; i = 1, 2, 3\}$, also called the detail subbands, convey the difference of information between r_{j-1} and the coarser approximation r_j (we refer to [34], [35] for a comprehensive introduction). By arranging r_J together with the w_j^i 's in a single column vector \tilde{x} , the basis transfer scheme can be implemented with an analysis and synthesis matrix pair (see, e.g., [36]): $\tilde{x} = \mathfrak{A}_J x$, $x = \mathfrak{S}_J \tilde{x}$. Consequently, a configuration x^* is a global minimum of U if and only if $\mathfrak{A}_J x^*$ is a global minimum of

$$\tilde{U}_J : \mathbb{R}^{|S|} \rightarrow \mathbb{R}, \quad \tilde{x} \mapsto U(\mathfrak{S}_J \tilde{x})$$

so that the minimization can be equivalently performed on the wavelet domain. Working with wavelet representations is fruitful as regards restriction of the state space. The HF wavelet coefficients lend themselves well to quantization and their amplitude range can be estimated on the basis of the knowledge of the maximum discontinuity amplitude to be found in the original image. Moreover, the spatial orientation of the detail subbands allows LB descriptions with spatially oriented neighborhoods. This gives rise to a finite DWT space $\hat{\Omega}_J^{|S|}$ with reduced cardinality which is rich enough to contain close approximations to most piecewise smooth scenes. The construction of $\hat{\Omega}_J^{|S|}$ is the subject of Section III-A and annealing on this subspace is discussed in Section III-B.

A. Restricted Wavelet Space

Without loss of generality, we assume that the support S of the true image x^0 is $M \times M$ with $2^{-J}M \in \mathbb{N} \setminus \{0, 1\}$. Let S_j

and S_j^i denote the $2^{-J}M \times 2^{-J}M$ and $2^{-j}M \times 2^{-j}M$ grids supporting the low resolution residual r_J and the HF subband w_j^i , respectively. Since there necessarily exists a closed interval $\Upsilon = [\Upsilon_{\min}, \Upsilon_{\max}]$ such that $x^0 \in \Upsilon^{|S|}$, the DWT space can be restricted to the set

$$\Upsilon^{|S_j|} \times \left(\prod_{j=1}^J \prod_{i=1}^3 (\Upsilon_j^i)^{|S_j^i|} \right)$$

where $\Upsilon_j^i = [-c_j^i, c_j^i]$ is defined by the maximum absolute value c_j^i of the wavelet coefficients in the detail subband w_j^i associated with $\text{DWT}(x^0)$. Our first goal is to estimate c_j^i . Assuming that S rests on a torus, w_j^i can be obtained by convolving x^0 with the impulse response of a 2-D high-pass filter and sub-sampling by 2^j along the rows and columns. Let us respectively denote the DWT analysis low-pass and high-pass filters by h and g . We have

$$w_j^i(m, n) = (x^0 \otimes_2 \mathcal{A}_j^i)(2^j m, 2^j n)$$

where \otimes_n is the n -D discrete convolution operator and

$$\mathcal{A}_j^i(m, n) = \begin{cases} h_{j-1}(m)g_{j-1}(n) & \text{if } i = 1, \\ g_{j-1}(m)h_{j-1}(n) & \text{if } i = 2, \\ g_{j-1}(m)g_{j-1}(n) & \text{if } i = 3 \end{cases} \quad (8)$$

with $h_j = h \otimes_1 (h \uparrow 2) \otimes_1 \dots \otimes_1 (h \uparrow 2^j)$ (the notation $h \uparrow 2^l$ stands for the signal obtained by inserting $2^l - 1$ zeros between each neighboring samples of h), $g_j = h_{j-1} \otimes_1 (g \uparrow 2^j)$ if $j \geq 1$ and $g_0 = g$. Let $\mathcal{S}(\mathcal{A}_j^i)$ be the support of \mathcal{A}_j^i . Obviously, c_j^i satisfies

$$\begin{aligned} c_j^i &\leq \max_{z \in \Upsilon^{|S(\mathcal{A}_j^i)|}} \left| \sum_{(m,n) \in \mathcal{S}(\mathcal{A}_j^i)} z(m, n) \mathcal{A}_j^i(m, n) \right| \quad (9) \\ &= (\Upsilon_{\max} - \Upsilon_{\min}) \sum_{(m,n) \in \mathcal{S}(\mathcal{A}_j^i)} (\mathcal{A}_j^i(m, n))^+ \end{aligned}$$

but this upper bound is far from being sharp when x^0 is piecewise smooth. Inasmuch as we are concerned with nearly piecewise planar images, it is fair to limit the domain $\Upsilon^{|S(\mathcal{A}_j^i)|}$ in (9) to configurations having a single first-order discontinuity in a vertical ($i = 1$), horizontal ($i = 2$) or diagonal ($i = 3$) direction. In other words, given the maximum discontinuity amplitude \mathcal{K} to be found in x^0 , we consider the three sets $\{z_{a,l,\mathcal{K}}^i; a \in \mathbb{R}, l \in \mathbb{Z}\}$, $i = 1, 2, 3$, of distributions depicted in Fig. 1, that is,

$$z_{a,l,\mathcal{K}}^1(m, n) = \begin{cases} (\mathcal{K} - a)(n - l) & \text{if } n \geq l, \\ a(l - n) & \text{otherwise} \end{cases} \quad (10)$$

$$z_{a,l,\mathcal{K}}^2(m, n) = z_{a,l,\mathcal{K}}^1(n, m) \quad (11)$$

and

$$z_{a,l,\mathcal{K}}^3(m, n) = \begin{cases} \frac{(\mathcal{K} - a)(m + n - 2l)}{\sqrt{2}} & \text{if } m + n \geq 2l, \\ \frac{a(2l - m - n)}{\sqrt{2}} & \text{otherwise.} \end{cases} \quad (12)$$

Our estimate for c_j^i then writes

$$\hat{c}_j^i(\mathcal{K}) = \max_{a \in \mathbb{R}, l \in \mathbb{Z}} \left| \sum_{m,n} z_{a,l,\mathcal{K}}^i(m, n) \mathcal{A}_j^i(m, n) \right|.$$

For $\sum_n h(n) = 1$ and assuming the mother wavelet has at least two vanishing moments, this gives

$$\hat{c}_j^1(\mathcal{K}) = \hat{c}_j^2(\mathcal{K}) = \mathcal{K} \max_{l \in \mathbb{Z}} \left| \sum_{m>l} (m-l) g_{j-1}(m) \right|$$

$$\text{and } \hat{c}_j^3(\mathcal{K}) = \sqrt{2} \mathcal{K} \max_{l \in \mathbb{Z}} \left| \sum_{m+n \geq 2l} g_{j-1}(m) g_{j-1}(n) \right|.$$

The next step toward the construction of the restricted DWT space is to quantize the wavelet coefficients. The motivation for discretization is twofold. First, positive convergence results dealing with the finite-time behavior of annealing are restricted to finite state spaces [32], [37]. Second, efficient selection of the final inverse temperature requires the computation of an isolated minimum through coordinate-wise deterministic minimization [26], which cannot be done on the Cartesian product of uncountably infinite sets and is far too time-consuming under floating-point approximation. We use linear scalar uniform quantization, although it is well-known that the statistical properties of the detail subband coefficients are adequately summarized by generalized Gaussian distributions [34], [38], [39]. The reason for this simple choice is that both blur and noise contaminate the necessary information for fixing the parameters of the generalized Gaussian model — estimating these parameters from the data actually presents many difficulties in exchange for uncertain results. Let \mathcal{Q}_J (resp., \mathcal{Q}_j^i) be the number of quantization bins associated with r_J (resp., w_j^i). The quantized DWT subspace is

$$\tilde{\Lambda}_J^{lS_J} = \tilde{\Lambda}^{lS_J} \times \left(\prod_{j=1}^J \prod_{i=1}^3 (\tilde{\Lambda}_j^i)^{|S_j^i|} \right)$$

$$\text{with } \tilde{\Lambda} = \left\{ \Upsilon_{\min} + (\Upsilon_{\max} - \Upsilon_{\min}) \frac{(l + \frac{1}{2})}{\mathcal{Q}_J}; \right. \\ \left. l = 0, \dots, \mathcal{Q}_J - 1 \right\}$$

$$\text{and } \tilde{\Lambda}_j^i = \left\{ \hat{c}_j^i(\mathcal{K}) \left(\frac{(2l+1)}{\mathcal{Q}_j^i} - 1 \right); l = 0, \dots, \mathcal{Q}_j^i - 1 \right\}.$$

Of course, the \mathcal{Q}_j^i 's have to be odd so that the detail subband coefficients can be equal to zero. In practice, we use the same number of quantization bins, say \mathcal{Q}_{DS} , for all the detail subbands: $\mathcal{Q}_j^i = \mathcal{Q}_{DS} \in \mathbb{N} \setminus \{0, 1\}$ for all $(i, j) \in \{1, 2, 3\} \times \{1, \dots, J\}$. It is also worthwhile noting that the expected maximum discontinuity amplitude can be over-estimated in order not to penalize possible abrupt intensity changes. We suggest to choose $\mathcal{K} \approx (\Upsilon_{\max} - \Upsilon_{\min})/2$, which corresponds to $\mathcal{K} \approx 128$ grey levels in standard 8-bit images. Then, from our experience, setting $\mathcal{Q}_J \geq 2^7$ and $\mathcal{Q}_{DS} = \mathcal{Q}_J - 1$ is appropriate to most applications.

An immediate refinement in this quantized subspace is to restrict the set $\tilde{\Lambda}^{lS_J}$ of low resolution residuals to an LB subset with level ζ_J and four-nearest neighbor system $\mathcal{N}(S_J)$ — this subset is denoted by $\Omega(\zeta_J, \tilde{\Lambda}, \mathcal{N}(S_J))$ in accordance with the notation introduced in Section II-B. Besides, keeping in mind that the detail subbands can be interpreted as spatially oriented HF channels, the most likely configurations in $(\tilde{\Lambda}_j^i)^{|S_j^i|}$ should be well-described by an LB subspace with suitably oriented

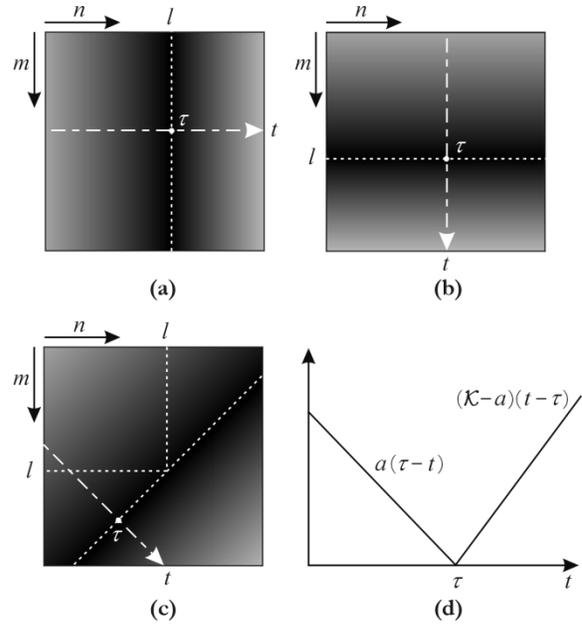


Fig. 1. Prototypical distributions $z_{a,l,\mathcal{K}}^i$ [see (10)–(12)] considered for the estimation of the amplitude ranges of the detail subbands: (a) $i = 1$; (b) $i = 2$; (c) $i = 3$; and (d) intensity change along the t -axis.

neighborhood system. Since w_j^1 gives the horizontal high frequencies (i.e., vertical edges), w_j^2 the vertical high frequencies (i.e., horizontal edges) and w_j^3 the high frequencies in both directions, it is natural to choose the 2-nearest vertical and horizontal neighbor systems for $i = 1$ and $i = 2$, respectively, and the 4-nearest diagonal neighbor system for $i = 3$. Hence, the restricted space we propose to consider is the set

$$\tilde{\Omega}_J^{lS_J} = \Omega(\zeta_J, \tilde{\Lambda}, \mathcal{N}(S_J)) \\ \times \left(\prod_{j=1}^J \prod_{i=1}^3 \Omega(\zeta_{DS}, \tilde{\Lambda}_j^i, \mathcal{N}^i(S_j^i)) \right) \quad (13)$$

where $\zeta_{DS} \in \{1, \dots, \mathcal{Q}_{DS} - 1\}$ denotes the LB approximation level associated with the w_j^i 's and

$$\mathcal{N}_{(m,n)}(S_J) = \{(m-1, n), (m+1, n), \\ (m, n-1), (m, n+1)\}, \\ \mathcal{N}_{(m,n)}^1(S_J^1) = \{(m-1, n), (m+1, n)\}, \\ \mathcal{N}_{(m,n)}^2(S_J^2) = \{(m, n-1), (m, n+1)\}, \\ \mathcal{N}_{(m,n)}^3(S_J^3) = \{(m-1, n-1), (m+1, n-1), \\ (m-1, n+1), (m+1, n+1)\}.$$

In a large number of experiments, we have found that $(\mathcal{Q}_J, \zeta_J, \mathcal{Q}_{DS}, \zeta_{DS}) = (128, 10, 127, 10)$ gives consistently good descriptions. In fact, the subspace $\tilde{\Omega}_J^{lS_J}$ defined by this set of parameter values is rich in the sense that for any reasonably smooth image $x \in \Upsilon^{lS_J}$ and for any $J \geq 1$ such that $2^{-J}M \gg 1$, there exists an x' whose DWT is in $\tilde{\Omega}_J^{lS_J}$ such that x and x' are not visually distinguishable. This is illustrated by the examples in Fig. 2. We used biorthogonal spline wavelets with two vanishing moments [40] and the restricted wavelet space approximations [Fig. 2(c) and (d)] of the original images displayed in Figs. 2(a) and (b) were obtained by estimating $\arg \min_{x \in \tilde{\Omega}_J^{lS_J}} \|\mathfrak{S}_J \tilde{x} - x^0\|_2^2$ using an annealing procedure.



Fig. 2. Restricted wavelet space approximation: (a)–(b) 256×256 original images in $[0, 255]^{[S]}$; (c)–(d) images with DWT in the set $\tilde{\Omega}_J^{[S]}$ (13) defined by $J = 2, \mathcal{K} = 128$ and $(Q_J, \zeta_J, Q_{DS}, \zeta_{DS}) = (128, 10, 127, 10)$.

B. Annealing on $\tilde{\Omega}_J^{[S]}$

The communication mechanism associated with the Metropolis dynamics working on $\tilde{\Omega}_J^{[S]}$ (13) is implemented in the following manner. A new candidate \tilde{y} is generated from the current configuration, say

$$\tilde{x} = \{r_J[\tilde{x}], w_j^i[\tilde{x}]; j = 1, \dots, J, i = 1, 2, 3\}$$

by first randomly selecting a site s from

$$S_J \cup \left(\bigcup_{j=1}^3 S_j^i \right)$$

and then setting $\tilde{y}_{\setminus s} = \tilde{x}_{\setminus s}$ and affecting to \tilde{y}_s a random value in the section of $\Omega(\zeta_J, \tilde{\Lambda}, \mathcal{N}(S_J))$ at $(r_J[\tilde{x}])_{\setminus s}$ if $s \in S_J$ or in the section of $\Omega(\zeta_{DS}, \tilde{\Lambda}_j^i, \mathcal{N}^i(S_j^i))$ at $(w_j^i[\tilde{x}])_{\setminus s}$ if $s \in S_j^i$. Note that the site selection step can be equivalently performed by randomly choosing a grid \tilde{S} from $\{S_J, S_j^i; j = 1, \dots, J, i = 1, 2, 3\}$ with probability 4^{-J} for S_J and 4^{-j} for any S_j^i and then picking a site s from \tilde{S} . Thus, the communication kernel \tilde{q}' on $\tilde{\Omega}_J^{[S]}$ can be defined by (see (14) at the bottom of the next page) for all $\tilde{x}, \tilde{y} \in \tilde{\Omega}_J^{[S]}$ such that $\tilde{x} \neq \tilde{y}$, where q' is given in (7). Since q' is symmetric and irreducible, we can see from (14) that so is \tilde{q}' and it follows that the theoretical convergence properties of annealing hold for $\mathbf{M}(\tilde{\Omega}_J^{[S]}, \tilde{U}_J, \tilde{q}', (\beta_n))$. In particular, there exists finite-time exponential cooling schedule $(\beta_n) = (\beta_0 \exp(\kappa n))_{1 \leq n \leq N}$ with initial inverse temperature β_0 independent of the horizon N that are “logarithmically almost optimal” [32]: for any positive real β_0 , one can choose κ as a decreasing function of N such that the logarithm of the convergence measure

$$M_\alpha(N)$$

$$= \sup_{\tilde{x} \in \tilde{\Omega}_J^{[S]}} P \left(\tilde{U}_J(X_N) \geq \min_{\tilde{z} \in \tilde{\Omega}_J^{[S]}} \tilde{U}_J(\tilde{z}) + \alpha \mid X_0 = \tilde{x} \right)$$

is asymptotically equivalent to the logarithm of the best achievable convergence rate (i.e., $\inf_{(\beta_n)} M_\alpha(N)$) for α small enough. Furthermore, it should be stressed that $\mathbf{M}(\tilde{\Omega}_J^{[S]}, \cdot, \tilde{q}', (\beta_n))$ is not limited to cost functions defined on the image space. Although we have not investigated this possibility, our algorithm is also applicable to any locally computable penalty defined on the wavelet domain. This includes, for example, stabilization schemes based on complexity penalties [41], Gaussian mixture models [42], or generalized Gaussian distributions [43].

In practice, the estimate of the true image is obtained by computing the inverse DWT of the output of $\mathbf{M}(\tilde{\Omega}_J^{[S]}, \tilde{U}_J, \tilde{q}', (\beta_0 \exp(\kappa n))_{1 \leq n \leq N})$, where the length N of the annealing chain is fixed by the available computing resources. The parameters β_0 and κ of the cooling schedule are selected in accordance with the following well-accepted criterions.

- Most transitions should be accepted at the beginning of the annealing process, that is, the initial inverse temperature β_0 should be such that

$$\sum_{\tilde{z} \in \tilde{\Omega}_J^{[S]} \setminus \{\tilde{x}\}} \tilde{P}_{\beta_0}(\tilde{x}, \tilde{z})$$

is close to one for almost all $\tilde{x} \in \tilde{\Omega}_J^{[S]}$, where \tilde{P}_β is the transition kernel (5) associated with the energy landscape $(\tilde{\Omega}_J^{[S]}, \tilde{U}_J, \tilde{q}')$.

- Conversely, given an isolated local minimum $\tilde{\mu}$ of $(\tilde{\Omega}_J^{[S]}, \tilde{U}_J, \tilde{q}')$ (i.e., $\tilde{\mu} \in \tilde{\Omega}_J^{[S]}$ and $\tilde{U}_J(\tilde{z}) > \tilde{U}_J(\tilde{\mu})$ for all \tilde{z} such that $\tilde{q}'(\tilde{\mu}, \tilde{z}) > 0$), the probability to escape from $\tilde{\mu}$ should be virtually equal to zero when the number of

iterations approaches N . In other words, the final inverse temperature $\beta_N = \beta_0 \exp(\kappa N)$ should be such that

$$\sum_{\tilde{z} \in \tilde{\Omega}_j^{S_1} \setminus \{\tilde{\mu}\}} \tilde{P}_{\beta_N}(\tilde{\mu}, \tilde{z})$$

is close to zero.

We refer to [26] for a description of efficient procedures to perform these estimation tasks. Explanations about the computation of the variation of the cost function $\tilde{U}_j^{(2)} : \tilde{x} \mapsto U^{(2)}(\mathfrak{S}_j \tilde{x})$ when changing a wavelet coefficient are provided in Appendix I, where we also discuss the complexity of the algorithm. It can be seen that if the size of the blurring kernel is a fixed fraction of the size of the image to be recovered, then the computational complexity of annealing on $\tilde{\Omega}_j^{S_1}$ is asymptotic to the computational complexity of annealing on the image space.

Our experimental results about piecewise planar image restoration using the above described stochastic process appear in Section V-A.

IV. WAVELET DOMAIN EDGE CONTINUATION

In many image restoration problems involving a first-order model, the exploitation of the fundamental smoothness features of the object boundaries can help to improve the quality of the results. This section describes an efficient way to do so by imposing simple penalties on the wavelet coefficients.

A. Method

We start with the common observation that the level curves of most images deserving interest are noiseless. From the wavelet domain standpoint, this prior information translates to the fact that the horizontal and vertical detail subbands are respectively vertically- and horizontally-oriented, that is, the vertical (resp. horizontal) gradient amplitudes values to be found in $w_j^1[\mathfrak{A}_j x^0]$ (resp. $w_j^2[\mathfrak{A}_j x^0]$) are expected to be small. Hence, the prolongation of discontinuities at different resolution levels can be facilitated in a natural way by means of an additional penalty term

$$\tilde{\Phi}(x) = \sum_{j=1, \dots, J} \sum_{l=1, 2} \tilde{\phi} \left(\frac{[\mathcal{D}_j^i w_j^i[\mathfrak{A}_j x]]_l}{\tilde{\Delta}} \right) \quad (15)$$

where the function $\tilde{\phi} : \mathbb{R} \rightarrow \mathbb{R}^+$ is even, increasing in \mathbb{R}_+ and such that $\tilde{\phi}(0) = 0$, \mathcal{D}_j^i is the matrix implementation of a gradient operator in a vertical ($i = 1$) or horizontal ($i = 2$) direction, and $\tilde{\Delta} \in \mathbb{R}_+^*$. The action of $\tilde{\Phi}$ is easy to understand in the frequency domain: it penalizes configurations that convey significant information in the regions schematized in Fig. 3(a) and exemplified in Fig. 3(b). This shows that $\tilde{\Phi}$ not only favors the formation of horizontal and vertical edges but also preserves diagonal boundaries. In addition, it avoids the introduction of an explicit line process (either Boolean [1], [29] or continuous

[44], [45]) and the associated increase in problem complexity is small if $\tilde{\phi}$ is chosen to be convex.

A guideline for selecting $\tilde{\phi}$ in (15) is that $\tilde{\Phi}$ should preserve both sharp transitions and smooth variations along the orientation directions of the wavelet subbands. By way of illustration, consider the contribution to $\tilde{\Phi}$ of any column n of a given horizontal subband w_j^1 , i.e.,

$$\sum_m \tilde{\phi} \left(\frac{(w_j^1[\mathfrak{A}_j x](m, n) - w_j^1[\mathfrak{A}_j x](m-1, n))}{\tilde{\Delta}} \right).$$

Our wish is that this contribution does not introduce any particular bias toward a discontinuous or smooth approximation to the original signal $w_j^1[\mathfrak{A}_j x^0](\cdot, n)$. Hence, according to comments made by Bouman and Sauer [13], we shall use $\tilde{\phi}(w) = |w|$. Our estimate of the original distribution x^0 is then defined as any global minimum of an augmented cost function of the form

$$V(x) = U^{(1)}(x) + \tilde{\lambda} \tilde{\Phi}(x) \\ = \|\mathcal{H}x - d\|_2^2 + \lambda \Phi^{(1)}(x) + \tilde{\lambda} \tilde{\Phi}(x) \quad (16)$$

where $\Phi^{(1)}$ (2)–(3) was specified in the introduction and $\tilde{\lambda} \in \mathbb{R}_+^*$ weights the influence of the wavelet smoothing term

$$\tilde{\Phi}(x) = \sum_{j=1, \dots, J} \sum_{l=1, 2} \left| [\mathcal{D}_j^i w_j^i[\mathfrak{A}_j x]]_l \right| \quad (17)$$

(the scale parameter $\tilde{\Delta}$ has been absorbed by $\tilde{\lambda}$).

From our experience, annealing on the wavelet domain and annealing on the image space exhibit similar performance in the case of a first-order prior model. Therefore, in practice, we suggest to minimize V by using $\mathbf{M}(\Omega^{S_1}, V, q', (\beta_n))$ as defined in Section II-B. Implementation details are given in Appendix I, where it is shown that, for large support blurring kernels, the extra processing cost associated with the treatment of (17) is a small fraction of the overall processing cost.

B. Parameter Selection

The hyper-parameters associated with $V(16)$ –(17) are λ , Δ , α , and $\tilde{\lambda}$. In practice, Δ is typically chosen to be of the order of the mean discontinuity amplitude to be recovered and the parameter α , which controls the sharpness of the restored edges, is fixed by the requirements of the application. Also, a useful lower bound for λ can be found in [5]. In truth, the main difficulty here is to select an appropriate value for $\tilde{\lambda}$. Immediate suggestions include the Chi-square choice (see, e.g., [46]), generalized cross-validation [47], or the L-curve method [48]. In our case, however, there is no clear justification for any of these approaches and, on top of this, the corresponding computational requirements would be very high. Instead, in a Bayesian framework, we could adopt a maximum likelihood estimation approach (see, e.g., [16], [49]) or we could construct an a priori distribution for λ and estimate it together with the original image

$$q'(\tilde{x}, \tilde{y}) = \begin{cases} 4^{-J} q' \left(r_J[\tilde{x}], r_J[\tilde{y}]; \zeta_J, \tilde{\Lambda}, \mathcal{N}(S_J) \right) & \text{if } r_J[\tilde{x}] \neq r_J[\tilde{y}], \\ 4^{-j} q' \left(w_j^i[\tilde{x}], w_j^i[\tilde{y}]; \zeta_{DS}, \tilde{\Lambda}_j^i, \mathcal{N}^i(S_j^i) \right) & \text{if } w_j^i[\tilde{x}] \neq w_j^i[\tilde{y}] \end{cases} \quad (14)$$

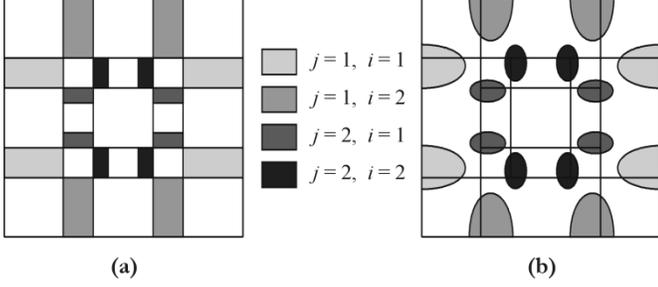


Fig. 3. (a) Frequency regions penalized by the wavelet smoothing term (15) and (b) 3 dB passbands of the filtering operations associated with $D_j^i w_j^i [2L_j]$ for biorthogonal spline wavelets with two vanishing moments.

using the maximum a posteriori criterion. Still, owing to the nonconvexity of $U^{(1)}$, both possibilities also present many computational difficulties. Alternatively, starting with some ideas of Geman and Reynolds [5], we have found a simple and efficient way to select $\tilde{\lambda}$ as a function of λ , Δ , α , \mathcal{H} , and the noise standard deviation σ . More specifically, Geman and Reynolds established a condition on λ in terms of Δ , \mathcal{H} and σ such that $U^{(1)}$ has the following property: if x^0 belongs to a certain class of prototypical images, then there is a high probability that x^0 is a coordinate-wise minimum of $U^{(1)}$, that is, a change of x^0 at anyone of the pixel values will result in an increase of $U^{(1)}$. For our part, we have chosen to seek a condition on $\tilde{\lambda}$ such that V satisfies that same property.

Let $\mathcal{L}^s \in \mathbb{R}^{|S|}$ be the configuration defined by $\mathcal{L}_t^s = \mathbf{1}_{\{t=s\}}$, $\mathcal{L} \in \mathbb{R}^*$. We say that $x \in \mathbb{R}^{|S|}$ is a coordinate-wise minimum of V if $(\forall s \in S)(\forall \mathcal{L} \in \mathbb{R}^*) [V(x + \mathcal{L}^s) > V(x)]$. Set

$$\delta_{s,\mathcal{L}} V(\eta) := V(x^0 + \mathcal{L}^s) - V(x^0) \quad (18)$$

where the argument η reminds us the dependence on the noise process, and let \mathcal{E} be the event that x^0 is a coordinate-wise minimum of V , i.e.,

$$\mathcal{E} = \{\eta \mid (\forall s \in S)(\forall \mathcal{L} \in \mathbb{R}^*) [\delta_{s,\mathcal{L}} V(\eta) > 0]\}. \quad (19)$$

Then, given λ , Δ , α , \mathcal{H} and σ , our concern is to find a constraint to place upon $\tilde{\lambda}$ in order to guarantee $P(\mathcal{E}) \geq 1 - \epsilon$, $\epsilon \in (0, 1)$, when x^0 is a member of a relevant image class. Following [5], we analyze the case of a vertical (or, equivalently, horizontal) step-edge, which amounts to assume that, for all m , $x^0(m, n) = \mathcal{K} \in \mathbb{R}_+^*$ if n is greater than some appointed value, 0 otherwise. We also suppose that \mathcal{H} is space-invariant, η is a Gaussian white noise, and S rests on a torus. Under these assumptions, we have the following theorem whose proof appears in Appendix II (the finite support signals h_j and g_j involved in the definition of the quantity ξ_J (22) are defined in Section III-A).

Theorem 1: Let $\epsilon \in (0, 1)$ and let χ be defined by

$$\text{erf}(\chi) := \frac{2}{\sqrt{\pi}} \int_0^\chi e^{-t^2} dt = (1 - \epsilon)^{1/|S|}. \quad (20)$$

Then, $P(\mathcal{E}) \geq 1 - \epsilon$ provided that (see (21) at the bottom of the page) where $\tau := \sum_s \mathcal{H}_s^2$ denotes the sum of squares of the blur coefficients and

$$\begin{aligned} \tilde{\xi}_J &:= \sum_{j=1, \dots, J} \tilde{\xi}_j^h \tilde{\xi}_j^g \\ \text{with } \tilde{\xi}_j^h &:= \min_{k \in \{0, \dots, 2^j - 1\}} \sum_n \left| h_{j-1}(2^j n + k) \right. \\ &\quad \left. - h_{j-1}(2^j(n-1) + k) \right| \\ \text{and } \tilde{\xi}_j^g &:= \min_{k \in \{0, \dots, 2^j - 1\}} \sum_n |g_{j-1}(2^j n + k)|. \end{aligned} \quad (22)$$

As we might expect, the lower bound $\tilde{\lambda}^*$ is an increasing function of σ and it decreases as λ increases. Moreover, setting $\tilde{\lambda} = 0$ in (21), we obtain

$$\lambda \geq \lambda^* := \begin{cases} \sqrt{2\tau} \Delta \sigma \chi & \text{if } \sigma \leq \frac{\sqrt{\tau} \Delta}{2\sqrt{2\alpha\chi}} \\ \frac{\tau \Delta^{1-\alpha}}{2\alpha} \left(\frac{\alpha}{1+\alpha} \left(2\sqrt{\frac{2}{\tau}} \sigma \chi + \Delta \right) \right)^{1+\alpha} & \text{otherwise} \end{cases} \quad (23)$$

and one can check that, if $\alpha = 1$, λ^* identifies with the bound computed by Geman and Reynolds [5, Theorem 2] in the first-order case (note that our parameter λ corresponds to λ^{-1} in their work and that their analysis lead to the same result for the four- and eight-nearest neighbor systems). Choosing $\tilde{\lambda} = \lambda^*$, we are free to select $\tilde{\lambda}$ as a function of λ . Yet it is important to be clear that the associated restoration results are then roughly equivalent to the ones obtained by the standard approach if λ is too close to λ^* , whereas the regions where the first-order differences are null cannot be recovered if λ is too small. We have found that values of λ ranging from $\lambda^*/8$ to $\lambda^*/2$ constitute a good balance. It should also be stressed that Theorem 1 stems from a worst-case analysis in the sense that any pixel is assumed to be adjacent to an edge. Consequently, χ (20) can be defined by a much smaller number than $|S|$, and our experiments in Section V-B confirm that setting $\chi = 3$ is sufficient.

V. EXPERIMENTS

Two sets of experimental results are presented here. One is for the wavelet domain image restoration approach described in Section III and the other concerns the edge continuation issue discussed in Section IV.

In every experiment, the data are obtained by degrading a given original image x^0 according to the linear observation model $d = \mathcal{H}x^0 + \eta$. The blur is either uniform or generated by convolving the 3×3 uniform mask with itself m times, which gives a $Q \times Q$ mask ($Q = 2m + 3$) with approximately Gaussian shape referred to as the ‘‘Gaussian $Q \times Q$ ’’ blur [5]. For the sake of clarity, the variance σ^2 of η is determined via

$$\tilde{\lambda} \geq \tilde{\lambda}^* := \begin{cases} \tilde{\xi}_J^{-1} (\sqrt{2\tau} \sigma \chi - \lambda \Delta^{-1}) & \text{if } \lambda \leq \frac{\tau \Delta^2}{2\alpha}, \\ \tilde{\xi}_J^{-1} \left(\sqrt{2\tau} \sigma \chi + \frac{\tau \Delta}{2} \left[1 - \left(1 + \frac{1}{\alpha} \right) \left(\frac{2\alpha\lambda}{\tau \Delta^2} \right)^{1/(1+\alpha)} \right] \right) & \text{otherwise} \end{cases} \quad (21)$$

the decibel level of the blurred signal-to-noise ratio (BSNR) as defined in [50]; that is, $(\text{BSNR})_{\text{dB}} = 10 \log_{10}(\sigma_e^2/\sigma^2)$, where σ_e^2 is the variance of the exact data $\mathcal{H}x^0$. We shall compare the performance of the proposed methods with the conventional annealing approach (see Section II-B), the iterative Wiener filtering technique described in [51], and the wavelet-based regularized deconvolution (WaRD) algorithm [52]. The metric used for assessing the quality of a solution x is the improvement in SNR (ISNR), given by $(\text{ISNR})_{\text{dB}} = 10 \log_{10}(\|x^0 - d\|_2^2/\|x^0 - x\|_2^2)$.

The annealing algorithms under examination operate either on an LB image subspace or on a restricted wavelet space, as discussed in Sections II-B and III-B. More specifically, we consider

$$\mathcal{A}(\cdot) := \mathbf{M}\left(\Omega^{|S|}, \cdot, q', (\beta_n)\right) \quad (24)$$

where $\Omega^{|S|}$ is defined by 256 quantization levels together with a four-nearest neighbor system and $\zeta = 5$ and

$$\tilde{\mathcal{A}}(\cdot) := \mathbf{M}\left(\tilde{\Omega}_J^{|S|}, \cdot, \tilde{q}', (\beta_n)\right) \quad (25)$$

with $\tilde{\Omega}_J^{|S|}$ (13) defined by the set of parameter values suggested in Section III-A. Both algorithms start from random configurations and use exponential cooling schedules with initial and final temperature values selected by means of the procedures proposed in [26]. Also, in all the experiments where the DWT is invoked, we use biorthogonal spline wavelets with two vanishing moments [40] and a wavelet representation on two resolution levels (i.e., $J = 2$).

A. Restoration in a Wavelet Basis

This section illustrates the benefits coming with the minimization of the cost function $U^{(2)}$ (1)–(3) over the wavelet domain. The annealing algorithms considered here are $\mathcal{A}(U^{(2)})$ (24) and $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ (25), where $\tilde{U}_J^{(2)} : \mathbb{R}^{|S|} \rightarrow \mathbb{R}$ is defined by $\tilde{U}_J^{(2)}(\tilde{x}) = U^{(2)}(\mathfrak{S}_J \tilde{x})$ for all \tilde{x} .

We begin with the restoration of the 128×128 piecewise planar image shown in Fig. 4(a) from the data displayed in Fig. 4(b). The original distribution has pixel intensity values in $[0, 36]$ and maximum first-order discontinuity amplitude equal to 3.5. The data were generated by blurring with the Gaussian 21×21 mask and adding 12 dB noise. Fig. 5 shows the restorations achieved in the dynamic range $[-15, 50]$ when running $\mathcal{A}(U^{(2)})$ and $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ for $8000 \cdot |S|$ iterations, together with the associated convergence rates. The prior was defined by $(\alpha, \Delta) = (1, 4)$ and the stabilization parameter λ was chosen according to the Geman and Reynolds' approach [5]. Because of the noise, the output of $\mathcal{A}(U^{(2)})$ has a mottled appearance. Clearly, $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ does not suffer from this weakness and produces markedly sharper discontinuities. Indeed, $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ offers substantial benefits in terms of both ISNR and final energy level ($U^{(2)} \approx 5.28 \cdot 10^4$ for the image in Fig. 5(a) and $U^{(2)} \approx 4.17 \cdot 10^4$ for Fig. 5(b)). Also, we can see from Table I that $256000 \cdot |S|$ iterations were not enough for $\mathcal{A}(U^{(2)})$ to produce a better solution than the one obtained with $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$

in only $8000 \cdot |S|$ iterations. From the material presented in Appendix I, for a $Q \times Q$ blurring kernel, the mean number of flops per iterations is lower than $3(Q^2 + 9(Q + 63))$ for $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ and equal to $3(Q^2 + 39)$ for $\mathcal{A}(U^{(2)})$. Here, the ratio of these two quantities is about 2.5 and it follows that the net computational gain brought by annealing on the wavelet domain is greater than $256/(8 \cdot 2.5) = 12.8$.

As a second example, consider the restoration problem depicted in Fig. 6(a) and (b). The original distribution, a 256×256 image with grey level range $[13\ 248]$, was degraded by uniform 11×11 blur and 30 dB noise. Restoration was first performed using iterative Wiener filtering [Fig. 6(c)] and the WaRD technique [Fig. 6(d)]. Figs. 7(a) and (b) display the restorations achieved with $\mathcal{A}(U^{(2)})$ in the dynamic range $[0, 255]$ for $(\alpha, \Delta) = (1, 4)$ and $(1, 8)$ (in both cases, the algorithm was run for $24000 \cdot |S|$ iterations and λ was again selected via the Geman and Reynolds' approach). We observe that annealing on the image space with a second-order prior can outperform the WaRD algorithm, but it also appears that its practical use is limited by sensitivity to the choice of the scale parameter. Running $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ under the same setting, we obtained the solutions shown in Figs. 7(c) and (d). Clearly, annealing on the wavelet domain is more stable than $\mathcal{A}(U^{(2)})$. Furthermore, it performs better than both $\mathcal{A}(U^{(2)})$ and WaRD in terms of ISNR.

B. Edge Continuation

In this section, we assess the improvements carried out by considering the augmented cost function V (16)–(17) rather than $U^{(1)}$. More specifically, we shall compare the performance of Wiener filtering, WaRD, $\mathcal{A}(U^{(1)})$, and $\mathcal{A}(V)$ (24). The annealing algorithms are always run for $8000 \cdot |S|$ iterations and χ is chosen to be 3 when computing the lower bounds $\tilde{\lambda}^*$ (21) and λ^* (23) for the wavelet smoothing parameter $\tilde{\lambda}$ and the stabilization parameter λ .

We start with the example depicted in Fig. 8. The original 256×256 image has pixel intensity values ranging from 0 to 255; it is observed through the Gaussian 9×9 blur and corrupted by 13 dB noise. Because of the high noise level, the Wiener filter and the WaRD algorithm perform poorly. Figs. 9(a)–(d) show the restorations achieved with $\mathcal{A}(U^{(1)})$ in the dynamic range $[0, 255]$ for $\alpha = 0.5$, $\Delta = 128$ and $\lambda = \lambda^*$, $\lambda^*/2$, $\lambda^*/4$ and $\lambda^*/8$, respectively ($\lambda^* = 1523$). Note that, compared to $\alpha = 1$, setting $\alpha = 0.5$ reduces the concavity of the ϕ -function (i.e., the amplitude of its second derivative) and hence the “patchiness” associated with first-order models. However, many discontinuity flaws are observed and increasing λ to correct these errors removes important details in the image. Moreover, the associated ISNR values show that $\mathcal{A}(U^{(1)})$ is outperformed by the WaRD algorithm in all four situations (though we do not necessarily find the WaRD restoration visually more attractive). Let us now introduce the wavelet smoothing term. Fig. 9(e) displays $\tilde{\lambda}^*$ as a function of λ (recall that $\tilde{\lambda}^* = 0$ for all $\lambda \geq \lambda^*$). The solutions obtained with $\mathcal{A}(V)$ for $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/2) = 16.1$, $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/4) = 26.1$ and $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/8) = 31.4$ are shown in Fig. 9(f)–(h). There are noticeable visual improvements: the boundaries are nearly

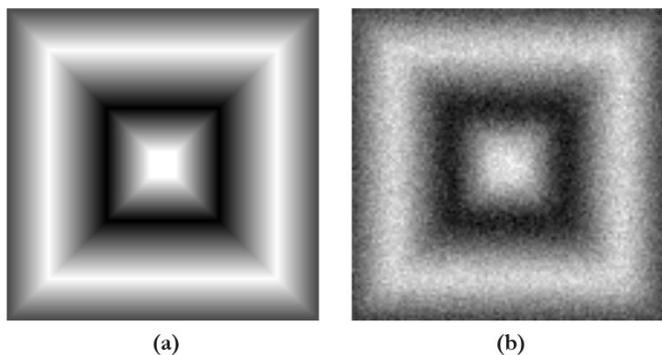


Fig. 4. Piecewise planar image: (a) original and (b) degraded observation: Gaussian 21×21 blur + 12 dB noise.

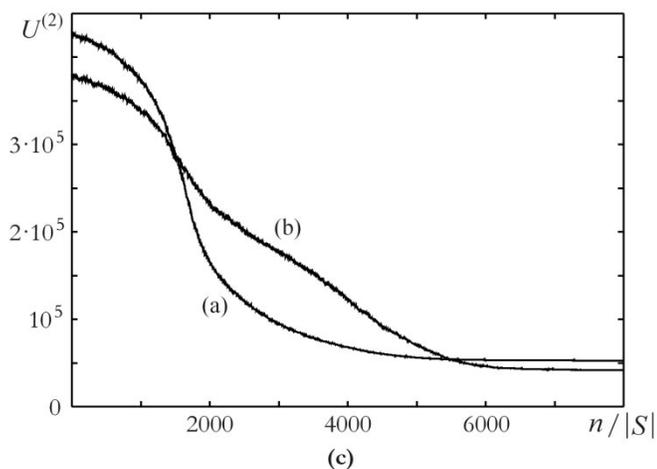
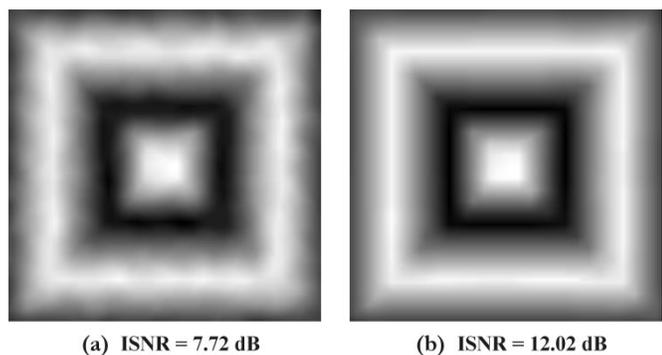


Fig. 5. Restoration results for the piecewise planar image (see Fig. 4): (a) annealing on the image space, (b) annealing on the wavelet domain, and (c) associated convergence rates.

flawless and the residual “patchiness” is unobtrusive. In addition, the corresponding ISNR values are 0.6 to 2 dB higher than for the output of the WaRD algorithm.

Similar experiments were conducted on the example show in Fig. 10. Figs. 11(a) and (b) display the restorations produced by $\mathcal{A}(U^{(1)})$ for $\lambda = \lambda^*/4$ and $\lambda = \lambda^*/8$ ($\lambda^* = 1113$). The results shown in Figs. 11(c) and (d) were obtained with $\mathcal{A}(V)$ for $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/4) = 20.8$ and $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/8) = 24.7$. Again, the additional wavelet smoothing term together with Theorem 1 yield better performance than WaRD. It should not be forgotten that our analysis for the selection of $\tilde{\lambda}$ is directed toward the formation of truly horizontal and vertical step-edges. Hence, for images with few horizontal and vertical boundaries such as

TABLE I
FINAL ENERGY LEVEL AND ISNR VERSUS LENGTH N OF THE ANNEALING CHAIN $\mathcal{A}(U^{(2)})$ WHEN APPLIED TO THE RESTORATION PROBLEM DEPICTED IN FIG. 4

$N/ S $	$U^{(2)}$	(ISNR) _{dB}
16000	$4.92 \cdot 10^4$	8.58
32000	$4.74 \cdot 10^4$	9.57
64000	$4.58 \cdot 10^4$	10.47
128000	$4.43 \cdot 10^4$	11.18
256000	$4.32 \cdot 10^4$	11.60
512000	$4.21 \cdot 10^4$	13.06



Fig. 6. Face image: (a) original; (b) degraded observation: uniform 11×11 blur + 30 dB noise; (c) restored by iterative Wiener filtering; and (d) restored with the WaRD algorithm.

the one considered here, we may ask ourselves whether further improvements can be achieved by diminishing $\tilde{\lambda}$. This turns out to be the case, as the results in Figs. 11(e) and (f) show, but the resulting benefits remain limited.

VI. CONCLUSION

Concave ϕ -functions stand out for boundary detection as well as homogeneous region recovery. Still, their practical use is limited by the fact that they generally lead to multimodal cost functions. Although simulated annealing is the method of choice to tackle this problem, standard stochastic restoration algorithms

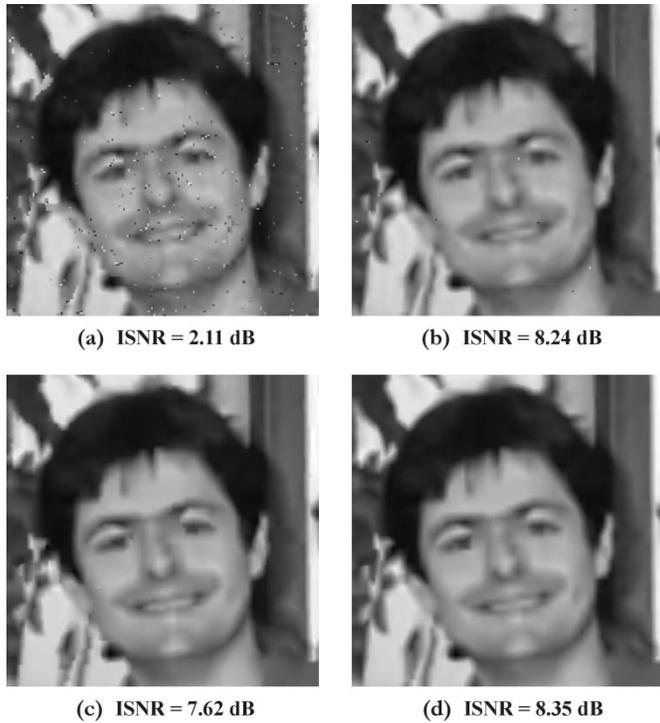


Fig. 7. Restoration results for the face image (see Fig. 6). (a) Annealing on the image space with $\Delta = 4$. Note that the outlier intensity values are at least two pixels wide because of the locally bounded description. (b) Annealing on the image space with $\Delta = 8$. (c)–(d) Annealing on the wavelet domain with $\Delta = 4$ and $\Delta = 8$, respectively.

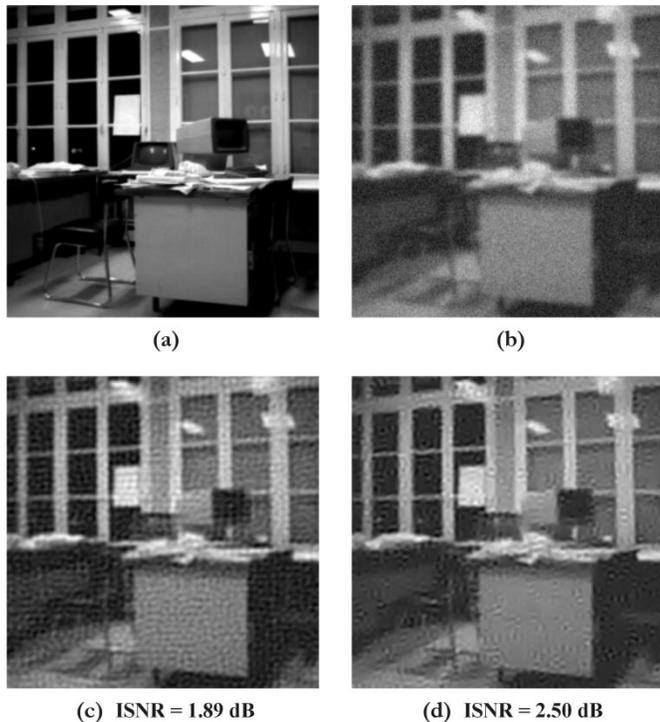


Fig. 8. Office image: (a) original; (b) degraded observation: Gaussian 9×9 blur + 30 dB noise; (c) restored by iterative Wiener filtering; and (d) restored with the WaRD algorithm.

have two weaknesses. First, single-site updating on the digital image space gets on badly with second-order models in the sense that it is very easy to be stuck in poor local minima.

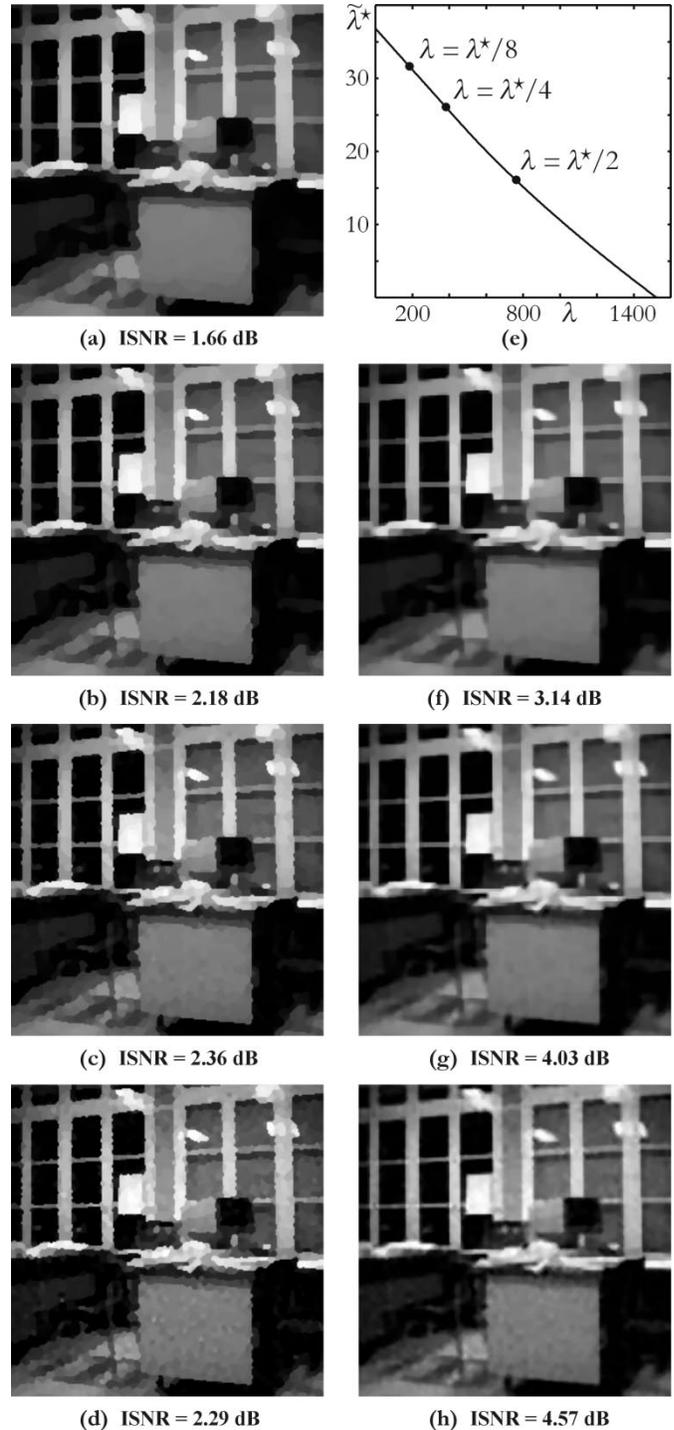


Fig. 9. Restoration results for the office image (see Fig. 8). (a)–(d) Minimization of $U^{(1)}$: (a) $\lambda = \lambda^*$; (b) $\lambda = \lambda^*/2$; (c) $\lambda = \lambda^*/4$; (d) $\lambda = \lambda^*/8$. (e) $\tilde{\lambda}^*(\lambda)$ curve (21). (f)–(h) Minimization of V : (f) $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/2)$; (g) $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/4)$; and (h) $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/8)$.

Second, the incorporation of the local geometry of edges via an explicit line field drastically increases computational complexity and requires ad hoc extra parameter specification. It has been shown that both difficulties can be overcome by resorting to the DWT.

We first demonstrated that the convergence toward global minima can be substantially improved by single-site updating on a suitably restricted wavelet space. Other “practical” versions

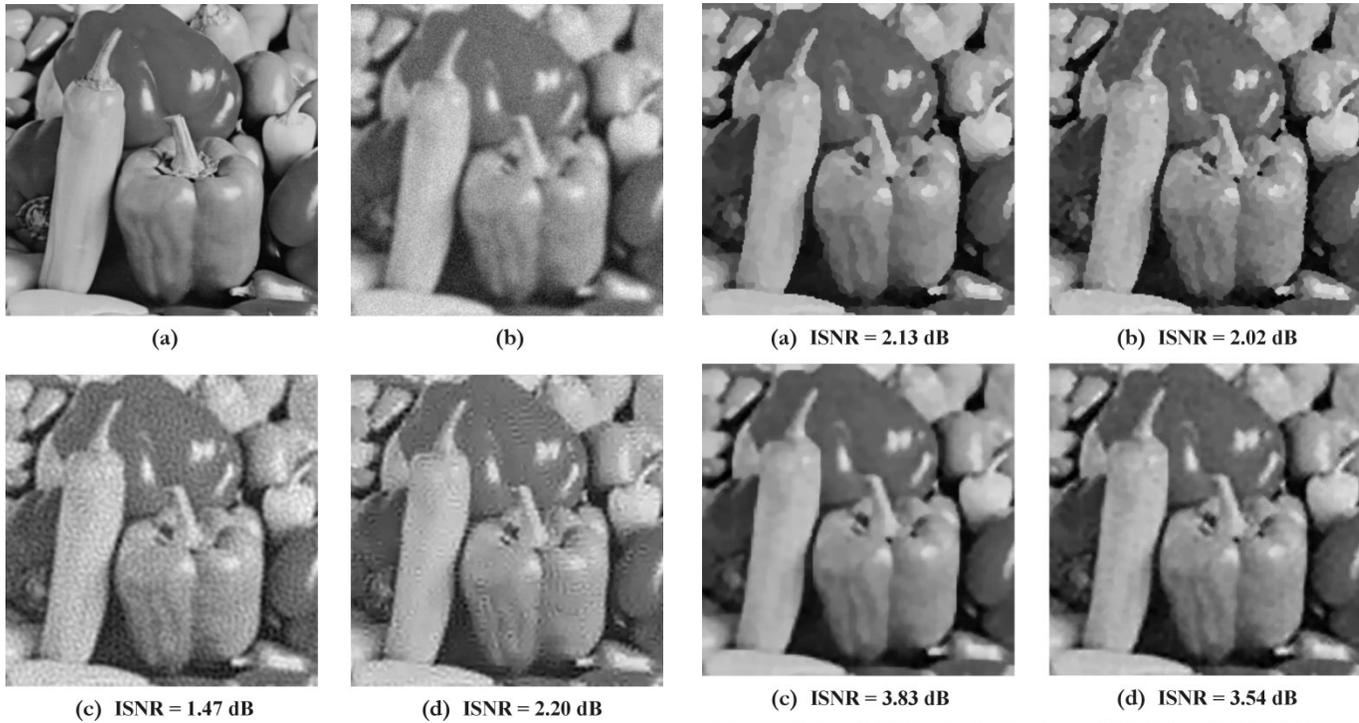


Fig. 10. Peppers image: (a) original; (b) degraded observation: Gaussian 9×9 blur + 13 dB noise; (c) restored by iterative Wiener filtering; and (d) restored with the WaRD algorithm.

of the proposed algorithm can be designed, but the merit of our approach is that the associated communication mechanism does not alter the theoretical convergence properties of annealing. We agree that our algorithm may seem complex in terms of the number of parameters defining the state space. However, it is not sensitive to moderate parameter variations and the suggested values produce good scene descriptions. Moreover, increasing the numbers of quantization bins together with the LB levels is usually useless because of the accuracy limitations imposed by the sensing environment.

We also introduced the idea of “implicitly interacting discontinuities” by means of an additional penalty term defined on the horizontal and vertical HF channels. In comparison with explicit line processes, the associated increase in problem complexity is small. Furthermore, we provided an explicit formula for choosing the extra parameter $\tilde{\lambda}$ as a function of the stabilization parameter λ , the blur coefficients, and the noise standard deviation. Using this result, our experiments show that boundary sharpness is preserved while the contour lines are smoother and hence more faithful to natural scenes. Such a behavior is highly desirable for subsequent feature extraction and segmentation tasks.

APPENDIX I IMPLEMENTATION AND COMPLEXITY OF THE PROPOSED ALGORITHMS

Recall that given an initial state $z_0 \in E$, a realization $(z_n)_{1 \leq n \leq N}$ of an annealing chain $\mathbf{M}(E, U, q, (\beta_n)_{1 \leq n \leq N})$ is generated as follows.

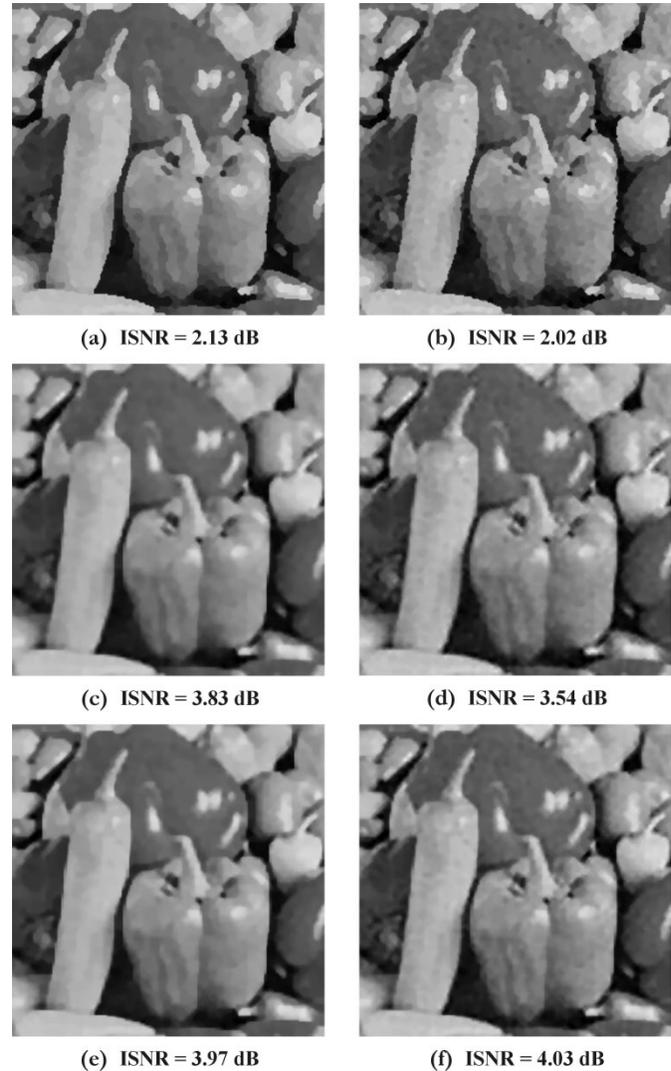


Fig. 11. Restoration results for the peppers image (see Fig. 10). (a)–(b) Minimization of $U^{(1)}$: (a) $\lambda = \lambda^*/4$; (b) $\lambda = \lambda^*/8$. (c)–(f) Minimization of \tilde{V} : (c) $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/4)$; (d) $\tilde{\lambda} = \tilde{\lambda}^*(\lambda^*/8)$; (e) $\tilde{\lambda} = 1/2\tilde{\lambda}^*(\lambda^*/4)$; and (f) $\tilde{\lambda} = 1/2\tilde{\lambda}^*(\lambda^*/8)$.

for $n = 1$ **to** N **do**

Propose a new state z' by drawing a sample from the distribution $q(z_{n-1}, \cdot)$ on E .

if $U(z') \leq U(z_{n-1})$ **then** Set $z_n = z'$.

else Set $z_n = z'$ with probability $\exp(-\beta_n(U(z') - U(z_{n-1})))$.

endfor

In image restoration applications, the computational complexity is governed by the evaluation of the energy difference $U(z') - U(z_{n-1})$ that takes place at each iteration. The following subsections provide some insights into the implementation of this critical task for annealing on the wavelet domain (see Section III-B) and for annealing on the image space with wavelet domain edge continuation (see Section IV). We drop the subscript n to simplify the notation and we denote by \mathcal{L}^s a configuration that is zero everywhere except at site s , where it

assumes the value \mathcal{L} . Consequently, z' is of the form $z + \mathcal{L}^s$ with \mathcal{L} in the section of the state space at $z_{\setminus s}$.

A. Annealing on the Wavelet Domain

The energy difference

$$\tilde{U}_J^{(2)}(z + \mathcal{L}^s) - \tilde{U}_J^{(2)}(z) =: \delta\tilde{U}_J^{(2)}$$

associated with $\mathbf{M}(\tilde{\Omega}_J^{|\mathcal{S}|}, \tilde{U}_J^{(2)}, \tilde{q}', (\beta_n)) =: \tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ can be written as

$$\begin{aligned} & \mathcal{L}^2 \sum_l [\mathcal{H}\mathfrak{G}_J]_{l,s}^2 + 2\mathcal{L} \sum_l [\mathcal{H}\mathfrak{G}_J]_{l,s} ([\mathcal{H}\mathfrak{G}_J]_l z - d_l) \\ & + \lambda \sum_l \left(\phi_\Delta \left([\mathcal{D}^{(2)}\mathfrak{G}_J]_l z + \mathcal{L} [\mathcal{D}^{(2)}\mathfrak{G}_J]_{l,s} \right) \right. \\ & \left. - \phi_\Delta \left([\mathcal{D}^{(2)}\mathfrak{G}_J]_l z \right) \right) \end{aligned}$$

where $[A]_l$ and $[A]_{l,s}$ respectively denote the l th row and the (l, s) entry of a matrix A and $\phi_\Delta(\cdot) := \phi(\cdot/\Delta)$. This suggests to precompute $\sum_l [\mathcal{H}\mathfrak{G}_J]_{l,s}^2 =: \theta_s$ for all s and to store the vectors $\mathcal{H}\mathfrak{G}_J z - d =: b$, $\mathcal{D}^{(2)}\mathfrak{G}_J z =: b'$, and $[\phi_\Delta(b'_1), \dots, \phi_\Delta(b'_{L_2})]^T =: b''$ into buffers. Thus,

$$\begin{aligned} \delta\tilde{U}_J^{(2)} &= \mathcal{L}^2 \theta_s + 2\mathcal{L} \sum_l [\mathcal{H}\mathfrak{G}_J]_{l,s} b_l \\ &+ \lambda \sum_l \left(\phi_\Delta \left(b'_l + \mathcal{L} [\mathcal{D}^{(2)}\mathfrak{G}_J]_{l,s} \right) - b''_l \right) \quad (26) \end{aligned}$$

and, in the event that the proposed move is accepted, b , b' and b'' have to be updated according to $b_l \leftarrow b_l + \mathcal{L}[\mathcal{H}\mathfrak{G}_J]_{l,s}$, $b'_l \leftarrow b'_l + \mathcal{L}[\mathcal{D}^{(2)}\mathfrak{G}_J]_{l,s}$ and $b''_l \leftarrow \phi_\Delta(b'_l)$ for all l . The number of flops involved in the computation of (26) or in the buffer updating procedure is linear in the numbers $n_s(\mathcal{H}\mathfrak{G}_J)$ and $n_s(\mathcal{D}^{(2)}\mathfrak{G}_J)$ of nonzero entries in columns s of $\mathcal{H}\mathfrak{G}_J$ and $\mathcal{D}^{(2)}\mathfrak{G}_J$. Assume for simplicity that the synthesis filters have the same number of taps, denoted by ρ . For a blurring kernel with discrete support of size $Q \times Q$, we have

$$\begin{aligned} n_s(\mathcal{H}\mathfrak{G}_J) &= \left(Q + \rho(2^{j(s)} - 1) - 2^{j(s)} + 1 \right)^2 \\ \text{and } n_s(\mathcal{D}^{(2)}\mathfrak{G}_J) &= 3 \left(\rho(2^{j(s)} - 1) - 2^{j(s)} + 3 \right)^2 - 2 \end{aligned}$$

where $j(s)$ stands for the decomposition level of the subband containing site s . Using these expressions, straightforward but tedious calculations show that the mean number of flops per iteration of $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ is

$$\begin{aligned} & 3(Q^2 + 4c_J Q + 88c_J + 48c'_J + 39), \\ \text{where } c_J &= (\rho - 1)(1 - 2^{-J}) \\ \text{and } c'_J &= (\rho - 1)^2 \left(\frac{3J}{4} + 2^{-J} - 1 \right). \end{aligned}$$

Setting $J = 0$ gives the mean number of flops per iteration of $\mathbf{M}(\Omega^{|\mathcal{S}|}, U^{(2)}, q', (\beta_n)) =: \mathcal{A}(U^{(2)})$ (i.e., annealing on the

image space), which is equal to $3(Q^2 + 39)$. Hence, for images of M^2 samples, both $\mathcal{A}(U^{(2)})$ and $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ have computational complexity of $O(\nu M^2 Q^2)$, where $\nu = N/M^2$ can be understood as the number of cycles through the pixels, or ‘‘sweeps’’.

Note that the storage requirements are high in the general case where the point spread function is space-variant. Still, in such a situation, the memory needs for $\tilde{\mathcal{A}}(\tilde{U}_J^{(2)})$ remain comparable to those for $\mathcal{A}(U^{(2)})$: using compressed row storage format with 4 bytes floating point numbers and 4 bytes integer indexes, $\mathcal{H}\mathfrak{G}_J$ and $\mathcal{D}^{(2)}\mathfrak{G}_J$ together occupy less than $8M^2(Q^2 + 4c_J Q + 24c_J + 16c'_J + 12)$ bytes of memory, while \mathcal{H} requires about $8M^2 Q^2$ bytes (there is no need to store $\mathcal{D}^{(2)}$ for annealing on the image space).

B. Wavelet Domain Edge Continuation

The energy difference $V(z + \mathcal{L}^s) - V(z)$ to be computed at each iteration of $\mathbf{M}(\Omega^{|\mathcal{S}|}, V, q', (\beta_n)) =: \mathcal{A}(V)$ is of the form $\delta U^{(1)} + \tilde{\lambda} \cdot \delta\tilde{\Phi}$ with

$$\begin{aligned} \delta U^{(1)} &= \mathcal{L}^2 \sum_l [\mathcal{H}]_{l,s}^2 + 2\mathcal{L} \sum_l [\mathcal{H}]_{l,s} ([\mathcal{H}]_l z - d_l) \\ &+ \lambda \sum_l \left(\phi_\Delta \left([\mathcal{D}^{(1)}]_l z + \mathcal{L} [\mathcal{D}^{(1)}]_{l,s} \right) \right. \\ &\quad \left. - \phi_\Delta \left([\mathcal{D}^{(1)}]_l z \right) \right) \\ \text{and } \delta\tilde{\Phi} &= \sum_l \left(\left| [\tilde{\mathcal{D}}\mathfrak{A}_J]_l z + \mathcal{L} [\tilde{\mathcal{D}}\mathfrak{A}_J]_{l,s} \right| - \left| [\tilde{\mathcal{D}}\mathfrak{A}_J]_l z \right| \right) \quad (27) \end{aligned}$$

where the matrix $\tilde{\mathcal{D}}$ is any vertical concatenation of the \mathcal{D}_j^i 's. Assuming that the analysis filters have the same number of taps, say ρ , it is easy to check that the number of nonzero entries in any column of $\tilde{\mathcal{D}}\mathfrak{A}_J$ does not exceed $2J\rho^2$. The treatment of $\delta U^{(1)}$ is similar to the treatment of $\delta\tilde{U}_J^{(2)}$ described in the previous subsection (simply set $J = 0$ and replace $\mathcal{D}^{(2)}$ with $\mathcal{D}^{(1)}$). Also, (27) suggests to store the vector $\tilde{\mathcal{D}}\mathfrak{A}_J z =: \tilde{b}$ into a buffer to be updated according to $\tilde{b}_l \leftarrow \tilde{b}_l + \mathcal{L}[\tilde{\mathcal{D}}\mathfrak{A}_J]_{l,s}$ for all l if the proposed move is accepted. It follows that the mean number of flops per iteration of $\mathcal{A}(V)$ is less than $3Q^2 + 14J\rho^2 + 48$, while in the absence of the wavelet smoothing term (i.e., if $\tilde{\lambda} = 0$), the mean number of flops per iteration is about $3Q^2 + 48$.

APPENDIX II PROOF OF THEOREM 1

Our proof is divided into two parts: we first compute a lower bound for $\delta_{s,\mathcal{L}} V(\eta)$ (18) (see proposition below) from which we then show that (21) is a sufficient condition to guarantee that the event $\mathcal{E}(19)$ has probability greater than $1 - \epsilon$.

Proposition: Let $Z(\eta) = \{Z_s(\eta); s \in \mathcal{S}\}$ be the zero-mean, Gaussian random process defined by $Z_s(\eta) = \sum_t \mathcal{H}_{t-s} \eta_t$. Then, for any $s \in \mathcal{S}$ and for any $\mathcal{L} \in \mathbb{R}^*$

$$\delta_{s,\mathcal{L}} V(\eta) \geq \tau \mathcal{L}^2 - 2\mathcal{L} Z_s(\eta) + 2\lambda \phi \left(\frac{\mathcal{L}}{\Delta} \right) + 2\tilde{\lambda} |\mathcal{L}| \tilde{\xi}_J.$$

Proof: Let us put $\phi_\Delta(\cdot) := \phi(\cdot/\Delta)$. The energy difference at site s writes

$$\begin{aligned} \delta_{s,\mathcal{L}}V(\eta) &= \delta_{s,\mathcal{L}}\mathcal{H}(\eta) + \lambda\delta_{s,\mathcal{L}}\Phi^{(1)} + \tilde{\lambda}\delta_{s,\mathcal{L}}\tilde{\Phi} \\ \text{where } \delta_{s,\mathcal{L}}\mathcal{H}(\eta) &= \|\mathcal{H}(x^0 + \mathcal{L}^s) - d\|_2^2 - \|\mathcal{H}x^0 - d\|_2^2, \\ \delta_{s,\mathcal{L}}\Phi^{(1)} &= \sum_l \left(\phi_\Delta \left(\left[\mathcal{D}^{(1)}(x^0 + \mathcal{L}^s) \right]_l \right) \right. \\ &\quad \left. - \phi_\Delta \left(\left[\mathcal{D}^{(1)}x^0 \right]_l \right) \right), \\ \text{and } \delta_{s,\mathcal{L}}\tilde{\Phi} &= \sum_{\substack{j=1,\dots,J \\ i=1,2}} \sum_l \left(\left| \left[\mathcal{D}_j^i w_j^i [\mathfrak{A}_J(x^0 + \mathcal{L}^s)] \right]_l \right| \right. \\ &\quad \left. - \left| \left[\mathcal{D}_j^i w_j^i [\mathfrak{A}_J x^0] \right]_l \right| \right). \end{aligned}$$

It has been demonstrated in [5] that

$$\delta_{s,\mathcal{L}}\mathcal{H}(\eta) = \tau\mathcal{L}^2 - 2\mathcal{L}Z_s(\eta).$$

We shall complete the proof by showing that

$$(i) \delta_{s,\mathcal{L}}\Phi^{(1)} \geq 2\phi_\Delta(\mathcal{L}) \text{ and } (ii) \delta_{s,\mathcal{L}}\tilde{\Phi} \geq 2|\mathcal{L}|\tilde{\xi}_J.$$

i) Recall that $\Phi^{(1)}$ (and hence $\delta_{s,\mathcal{L}}\Phi^{(1)}$) is a summation over two-sites cliques associated with the eight-nearest neighbor system and that x^0 is assumed to be a vertical or horizontal step-edge of size \mathcal{K} . If s is not adjacent to the edge, then

$$\delta_{s,\mathcal{L}}\Phi^{(1)} = 4(\phi_\Delta(\mathcal{L}) + \phi_{\Delta\sqrt{2}}(\mathcal{L})) > 2\phi_\Delta(\mathcal{L})$$

and we do not have to go any further. If s is adjacent to the edge, then

$$\begin{aligned} \delta_{s,\mathcal{L}}\Phi^{(1)} &= 3\phi_\Delta(\mathcal{L}) + \phi_\Delta(\mathcal{K} \pm \mathcal{L}) - \phi_\Delta(\mathcal{K}) \\ &\quad + 2(\phi_{\Delta\sqrt{2}}(\mathcal{L}) + \phi_{\Delta\sqrt{2}}(\mathcal{K} \pm \mathcal{L}) - \phi_{\Delta\sqrt{2}}(\mathcal{K})). \end{aligned} \quad (28)$$

Since ϕ is even, increasing in \mathbb{R}_+ and strictly concave in \mathbb{R}_+ , we have, for all $(a, b) \in \mathbb{R}^2$,

$$\begin{aligned} \phi(a+b) &= \phi(|a+b|) \geq \phi(|a| - |b|) \\ &= \phi(\max\{|a|, |b|\} - \min\{|a|, |b|\}) \\ &\geq \phi(\max\{|a|, |b|\}) - \phi(\min\{|a|, |b|\}) \\ &\geq \phi(b) - \phi(a). \end{aligned}$$

Applying this inequality to (28) gives $\delta_{s,\mathcal{L}}\Phi^{(1)} \geq 2\phi_\Delta(\mathcal{L})$.

ii) Since x^0 is a vertical or horizontal step-edge, we have $\mathcal{D}_j^i w_j^i [\mathfrak{A}_J x^0] = 0$ for all $(i, j) \in \{1, 2\} \times \{1, \dots, J\}$ and it follows that

$$\delta_{s,\mathcal{L}}\tilde{\Phi} = 2|\mathcal{L}| \sum_{j=1,\dots,J} \sum_l \left| \left[\mathcal{D}_j^1 w_j^1 [\mathfrak{A}_J 1^s] \right]_l \right|.$$

Let \mathcal{A}_j^1 be the 2-D impulse response defined by (8). We have

$$\begin{aligned} w_j^1 [\mathfrak{A}_J 1^s](m, n) &= (1^s \otimes_2 \mathcal{A}_j^1)(2^j m, 2^j n) \\ &= \mathcal{A}_j^1(2^j m + k, 2^j n + l) \end{aligned}$$

where the integers k and l are fixed by the position of site s . Therefore,

$$\begin{aligned} \sum_l \left| \left[\mathcal{D}_j^1 w_j^1 [\mathfrak{A}_J 1^s] \right]_l \right| &= \sum_{m,n} \left| \mathcal{A}_j^1(2^j m + k, 2^j n + l) \right. \\ &\quad \left. - \mathcal{A}_j^1(2^j(m-1) + k, 2^j n + l) \right| \\ &= \sum_m \left| h_{j-1}(2^j m + k) - h_{j-1}(2^j(m-1) + k) \right| \\ &\quad \cdot \sum_n |g_{j-1}(2^j n + l)| \\ &\geq \tilde{\xi}_j^h \tilde{\xi}_j^g \end{aligned}$$

so that $\delta_{s,\mathcal{L}}\tilde{\Phi} \geq 2|\mathcal{L}|\tilde{\xi}_J$. ■

Let $F_{\lambda, \tilde{\lambda}} : \mathbb{R}^* \rightarrow \mathbb{R}$ be defined by

$$F_{\lambda, \tilde{\lambda}}(\mathcal{L}) = \frac{\tau\mathcal{L}}{2} + \frac{\lambda\phi_\Delta(\mathcal{L})}{\mathcal{L}} + \tilde{\lambda}\tilde{\xi}_J$$

for all $\mathcal{L} \in \mathbb{R}^*$. From the above proposition, we have

$$\mathcal{E} \supset \left\{ \eta \mid (\forall s \in S) (\forall \mathcal{L} \in \mathbb{R}_+^*) \left[F_{\lambda, \tilde{\lambda}}(\mathcal{L}) > |Z_s(\eta)| \right] \right\}$$

and, hence,

$$\mathbb{P}(\mathcal{E}) \geq \mathbb{P} \left(\bigcap_{s \in S} \left\{ |Z_s(\eta)| < R(\lambda, \tilde{\lambda}) \right\} \right)$$

where $R(\lambda, \tilde{\lambda}) := \inf \{ F_{\lambda, \tilde{\lambda}}(\mathcal{L}); \mathcal{L} \in \mathbb{R}_+^* \}$. Since any zero-mean, Gaussian random process $\{Z_i; i = 1, \dots, n\}$ satisfies

$$\mathbb{P} \left(\bigcap_{i=1}^n \{|Z_i| \leq a_i\} \right) \geq \prod_{i=1}^n \mathbb{P}(|Z_i| \leq a_i)$$

for all $(a_1, \dots, a_n) \in (\mathbb{R}_+^*)^n$ (see, e.g., [53]), it follows that

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &\geq \prod_{s \in S} \mathbb{P} \left(|Z_s(\eta)| < R(\lambda, \tilde{\lambda}) \right) \\ &= \left(\frac{2}{\sqrt{\pi}} \int_0^{R(\lambda, \tilde{\lambda})} \frac{1}{\sqrt{2\tau\sigma}} \exp\left(\frac{-v^2}{2\tau\sigma^2}\right) dv \right)^{|S|} \\ &= \left(\operatorname{erf} \left(\frac{R(\lambda, \tilde{\lambda})}{\sqrt{2\tau\sigma}} \right) \right)^{|S|}. \end{aligned}$$

Consequently, according to the definition of χ (20)

$$R(\lambda, \tilde{\lambda}) \geq \sqrt{2\tau\sigma}\chi \quad (29)$$

implies $\mathbb{P}(\mathcal{E}) \geq 1 - \epsilon$. For the function ϕ in (3), the derivative of $F_{\lambda, \tilde{\lambda}}$ is

$$F'_{\lambda, \tilde{\lambda}}(\mathcal{L}) = \frac{\tau}{2} - \frac{\lambda\alpha}{\Delta^2 \left(1 + \frac{|\mathcal{L}|}{\Delta}\right)^{1+\alpha}}, \quad \mathcal{L} \in \mathbb{R}^*.$$

Clearly, $F'_{\lambda, \tilde{\lambda}}$ is increasing in \mathbb{R}_+^* so that if $F'_{\lambda, \tilde{\lambda}}(0^+) \geq 0$, then $R(\lambda, \tilde{\lambda}) = F_{\lambda, \tilde{\lambda}}(0^+)$, whereas if $F'_{\lambda, \tilde{\lambda}}(0^+) < 0$, then $R(\lambda, \tilde{\lambda}) =$

$F'_{\lambda, \tilde{\lambda}}(\mathcal{L}_0)$, where \mathcal{L}_0 is the (unique) zero of $F'_{\lambda, \tilde{\lambda}}$ in \mathbb{R}_+^* . More specifically, see (30) at the bottom of the page. By substituting (30) into (29) we obtain (21), which completes the proof of the theorem.

REFERENCES

- [1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 721–741, 1984.
- [2] J. Besag, "On the statistical analysis of dirty pictures (with discussion)," *J. R. Statist. Soc. Ser. B*, vol. 48, no. 3, pp. 259–302, 1986.
- [3] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Statist. Assoc.*, vol. 82, no. 397, pp. 76–89, 1987.
- [4] M. Bertero, T. A. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, vol. 76, pp. 869–889, Aug. 1988.
- [5] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 3, pp. 367–383, 1992.
- [6] R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity preserving regularization of inverse visual problems," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 3, pp. 455–469, 1994.
- [7] S. Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 6, pp. 576–586, 1995.
- [8] M. Nikolova, "Regularization functions and estimators," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Lausanne, Switzerland, September 1996, pp. 457–460.
- [9] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, pp. 298–311, Feb. 1997.
- [10] S. Z. Li, "Close-form solution and parameter selection for convex minimization-based edge-preserving smoothing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 9, pp. 916–932, 1998.
- [11] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, 1990.
- [12] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–446, 1990.
- [13] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Processing*, vol. 2, no. 3, pp. 296–310, 1993.
- [14] H. R. Künsch, "Robust priors for smoothing and image restoration," *Ann. Inst. Statist. Math.*, vol. 46, no. 1, pp. 1–19, 1994.
- [15] S. Z. Li, Y. H. Huang, and J. S. Fu, "Convex MRF potential functions," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Washington, DC, Oct. 1995, pp. 296–299.
- [16] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Stat. Inst.*, vol. 52, pp. 5–21, 1987.
- [17] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 194–202, 1989.
- [18] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing*, vol. 4, pp. 932–946, July 1995.
- [19] M. Nikolova, "Markovian reconstruction using a GNC approach," *IEEE Trans. Image Processing*, vol. 8, pp. 1204–1220, Sept. 1999.
- [20] —, "Local strong homogeneity of a regularized estimator," *SIAM J. Appl. Math.*, vol. 61, no. 2, pp. 633–658, 2000.
- [21] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.
- [22] M. Nikolova, J. Idier, and A. Mohammad-Djafari, "Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF," *IEEE Trans. Image Processing*, vol. 7, pp. 571–585, Apr. 1998.
- [23] B. Hajek, "Cooling schedules for optimal annealing," *Math. Oper. Res.*, vol. 13, no. 2, pp. 311–329, 1988.
- [24] T.-S. Chiang and Y. Chow, "On the convergence rate of annealing processes," *SIAM J. Control Optim.*, vol. 26, no. 6, pp. 1455–1470, 1988.
- [25] O. Catoni, "Large deviations and cooling schedules for the simulated annealing algorithm" (in French), *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 307, pp. 535–538, 1988.
- [26] M. C. Robini, T. Rastello, and I. E. Magnin, "Simulated annealing, acceleration techniques and image restoration," *IEEE Trans. Image Processing*, vol. 8, pp. 1374–1387, Oct. 1999.
- [27] "Discussion on the meeting on the Gibbs sampler and other Markov chain Monte Carlo methods," *J. R. Statist. Soc. Ser. B*, vol. 55, no. 1, pp. 53–102, 1993.
- [28] S. Geman, D. E. McClure, and D. Geman, "A nonlinear filter for film restoration and other problems in image processing," *CVGIP: Graph. Models Image Process.*, vol. 54, no. 4, pp. 281–289, 1992.
- [29] L. Bedini, L. Benvenuti, E. Salerno, and A. Tonazzini, "A mixed-annealing algorithm for edge preserving image reconstruction using a limited number of projections," *Signal Process.*, vol. 32, no. 3, pp. 397–408, 1993.
- [30] B. Gidas, "A renormalization group approach to image processing problems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 2, pp. 164–180, 1989.
- [31] A. D. Sokal, "Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms," Cours de troisième cycle de la physique en Suisse romande, Lausanne, Switzerland, 1989.
- [32] O. Catoni, "Rough large deviation estimates for simulated annealing: Application to exponential schedules," *Ann. Probab.*, vol. 20, no. 3, pp. 1109–1146, 1992.
- [33] C. Yang, "Efficient stochastic algorithms on locally bounded image space," *CVGIP: Graph. Models Image Process.*, vol. 55, no. 6, pp. 494–506, 1993.
- [34] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, 1989.
- [35] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
- [36] M.-H. Yaou and W.-T. Chang, "Fast surface interpolation using multiresolution wavelet transform," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 7, pp. 673–688, 1994.
- [37] O. Catoni, "Rates of convergence for sequential annealing: A large deviation approach," in *Simulated Annealing: Parallelization Techniques*, R. Azencott, Ed. New York: Wiley, 1992, pp. 25–35.
- [38] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.
- [39] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Processing*, vol. 9, pp. 1532–1546, Sept. 2000.
- [40] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 45, no. 5, pp. 485–560, 1992.
- [41] J. Liu and P. Moulin, "Complexity-regularized image restoration," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Chicago, IL, Oct. 1998, pp. 555–559.
- [42] Y. Wan and R. D. Nowak, "Bayesian multiscale approach to joint image restoration and edge detection," *Proc. SPIE*, vol. 3813, pp. 73–84, July 1999.
- [43] M. Belge, M. E. Kilmer, and E. L. Miller, "Wavelet domain image restoration with adaptive edge-preserving regularization," *IEEE Trans. Image Processing*, vol. 9, pp. 597–608, Apr. 2000.
- [44] L. Blanc-Féraud, S. Teboul, G. Aubert, and M. Barlaud, "Nonlinear regularization using constrained edges in image reconstruction," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Lausanne, Switzerland, Sept. 1996, pp. 449–452.

$$R(\lambda, \tilde{\lambda}) = \begin{cases} \lambda \Delta^{-1} + \tilde{\lambda} \tilde{\xi}_J & \text{if } \lambda \leq \frac{\tau \Delta^2}{2\alpha}, \\ \frac{\tau \Delta}{2} \left[\left(1 + \frac{1}{\alpha}\right) \left(\frac{2\alpha \lambda}{\tau \Delta^2}\right)^{1/(1+\alpha)} - 1 \right] + \tilde{\lambda} \tilde{\xi}_J & \text{otherwise} \end{cases} \quad (30)$$

- [45] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Processing*, vol. 10, pp. 1001–1009, July 2001.
- [46] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterton, "A study of methods of choosing the smoothing parameter in image restoration by regularization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 4, pp. 326–339, 1991.
- [47] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.
- [48] C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.
- [49] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2024–2036, Dec. 1989.
- [50] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Mag.*, vol. 14, no. 2, pp. 24–41, 1997.
- [51] A. D. Hillery and R. T. Chin, "Iterative wiener filters for image restoration," *IEEE Trans. Signal Processing*, vol. 39, pp. 1892–1899, Aug. 1991.
- [52] R. Neelamani, H. Choi, and R. Baraniuk, "Wavelet-based deconvolution using optimally regularized inversion for ill-conditioned systems," *Proc. SPIE*, vol. 3813, pp. 58–72, July 1999.
- [53] Y. L. Tong, *The Multivariate Normal Distribution*. New York: Springer-Verlag, 1990.



Marc C. Robini (M'99) received the M.Sc. degree in electrical and computer engineering and the Ph.D. degree in signal and image processing from INSA-Lyon Scientific and Technical University, France, in 1993 and 1998, respectively.

In 1999, he joined the INSA-Lyon where he is currently a Research Associate at the CREATIS (research and applications center for signal and image processing) and an Associate Professor at the Department of Electrical and Computer Engineering and at the "First Cycle" Department. His current research interests include inverse problems, multiresolution image processing, and dynamic Monte Carlo methods.



Isabelle E. Magnin (M'85) received the Engineering Degree in 1977 and the "Doctorat d'état ès Sciences" in 1987 from INSA-Lyon Scientific and Technical University, France.

Since 1982, she has been a Researcher with the French National Institute for Health and Medical Research (INSERM). She is the Head of the CREATIS laboratory (a CNRS research unit) and she has more than 100 publications. Her main interest concerns medical image processing, dynamic imaging, and inverse problems.