Simulated Annealing Based Shift-Variant Image Restoration Using the Multiresolution Wavelet Transform

M. C. Robini and I. E. Magnin

CREATIS, CNRS research unit (UMR 5515), INSA 502, 69621 Villeurbanne cedex, France
Tel: +33 4 72 43 87 85, Fax: +33 4 72 43 85 26, e-mail: marc.robini@creatis.insa-lyon.fr

Abstract
We address the linear image restoration problem in the case of a spatially varying blur. Most of the time, the recovery of the original image from its degraded measurements is an ill-conditioned, underdetermined inverse problem. Here, stabilization is achieved via concave potential functions and minimization is carried out using Metropolis-type simulated annealing. Still, the ordinary approach can be subject to some convergence difficulties and it remains an ambitious challenge to take into account the mutual dependence between neighboring discontinuities. We first propose to improve convergence towards global minima through single-site updating in the spatial frequency space. For this purpose, a suitable restricted DWT space is introduced and it turns out that the resulting class of algorithms shows less sensitivity to the choice of the hyperparameters. Next, we show that the smoothness of the discontinuity field can be incorporated implicitly in a multiresolution framework by means of a simple penalty term defined on the high frequency channels.

1 Introduction
The discrete linear image restoration problem is to recover an original intensity distribution $x^0 = [x^0_1, \ldots, x^0_2]^T$, defined over a 2-D pixel grid $\mathcal{S}$, from the measurement vector

$$d = [d_1, \ldots, d_S]^T = \mathcal{H} x^0 + \eta, \quad S' < S,$$  \hspace{2em} (1)

where the matrix $\mathcal{H}$ represents a shift-variant blur operator and $\eta$ is generally assumed to be a white Gaussian noise process. In order to overcome the difficulties resulting from the ill-posedness of the initial continuous problem, most image restoration methods include some $a$ priori constraints about the true image $x^0$ in addition to those implicit in coherence to the data. A common estimate of $x^0$ is $\hat{x}^0 = \arg\min_{x \in \Lambda} U(x)$, where the cost functional $U : \Lambda^S \rightarrow \mathbb{R}, \quad \Lambda = [\Lambda_{\text{min}}, \Lambda_{\text{max}}] \subset \mathbb{R}$, is defined by

$$U(x) = (\mathcal{H} x - d)^T (\mathcal{H} x - d) + \lambda_h \Phi^{(h)}(x).$$  \hspace{2em} (2)

Here, $\lambda_h \in \mathbb{R}^+$ is the usual “smoothing parameter” and the “regularization term” $\Phi^{(h)}$ should be beneficial for the removal of blur and for the recovery of $k$th-order discontinuities.

We consider the model

$$\Phi^{(h)}(x) = \sum_{l=1}^{L_k} \phi \left( \frac{[D^{(h)} x]_l}{\Delta_k} \right),$$  \hspace{2em} (3)

introduced in [1], where $\Delta_k \in \mathbb{R}^+$ is a scale parameter, $D^{(h)} \in \mathbb{R}^{L_k \times S}$ is a $k$th-order discrete derivative operator and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is taken to be even, increasing in $\mathbb{R}_+$ and strictly concave in $\mathbb{R}^+$.

However, the above approach suffers from two drawbacks. Firstly, it is rather well known that Markov chain Monte Carlo algorithms defined on a large state space with loosely coupled components experience difficulties if these are updated singly (see, e.g., Jennison’s discussion in [5]). Secondly, the mutual dependence between neighboring discontinuities cannot be taken into account unless an explicit “line process” is introduced [6], which results in a severe increase in computational complexity.

In the present study, after a brief description of annealing algorithms, we show that both problems can be elegantly alleviated in a multiresolution scheme using the 2-D discrete wavelet transform (DWT). Since single-site updating in the spatial frequency space corresponds to some block-site updating in the spatial domain, larger component connections are achieved by minimizing $U$ with respect to $\hat{x} \equiv \text{DWT}(x)$. As emphasized in section 3, this leads to significant benefits.
as soon as the true image exhibits a strong low frequency content. Next, as discussed in section 4, we propose to introduce prior constraints about the interactions between discontinuities by means of an additional penalty term defined on the horizontal and vertical high frequency channels.

2 Annealing algorithms

Let us define an energy landscape \((E, V, q)\) where \(E\) is a finite set called the configuration space, or state space, \(V : E \rightarrow \mathbb{R}\) is any function to be minimized and \(q : E \times E \rightarrow [0, 1]\) is a symmetric and irreducible Markov matrix which we will call the communication kernel. Usually, one defines a neighborhood system \(\mathcal{N}^E = \{\mathcal{N}^E_x; x \in E\}\) on \(E\) and set

\[
q_{x,y} = \begin{cases} 
|A|^\frac{1}{-1} & \text{if } y \in \mathcal{N}^E_x, \\
0 & \text{otherwise},
\end{cases} \tag{4}
\]

where \(|A| = \text{card } A\). For instance, in typical applications of the Metropolis dynamics to image restoration, \(E = \Delta^S\), \(\Delta = \{0, \ldots, q-1\}\), and

\[
\mathcal{N}^E_x = \{ y \in E \mid \{ s \mid x_s \neq y_s \} = 1 \}. \tag{5}
\]

An annealing algorithm on \((E, V, q)\) is a discrete-time, nonhomogeneous Markov chain \((X_n)_{n \in \mathbb{N}}\) on \(E\) specified by \((E, V, q, \nu_0, (\beta_n)_{n \in \mathbb{N}^*})\), where \(\nu_0\) is the initial law of \(X_0\) and \((\beta_n)_{n \in \mathbb{N}^*}\) is an increasing sequence of positive parameters (inverse temperatures) called the cooling schedule. The transitions \(P(X_n = y \mid X_{n-1} = x) = p_{x,y}^{V, \beta_n}\) of \((X_n)_{n \in \mathbb{N}}\) are expressed in the following way:

\[
p_{x,y}^{V, \beta_n} = q_{x,y} \exp (-\beta_n (V_y - V_x)^+), \quad \text{if } x \neq y, \tag{6}
\]

\[
p_{x,x}^{V, \beta_n} = 1 - \sum_{z \in E \setminus \{x\}} p_{x,z}^{V, \beta_n}. \tag{7}
\]

where \(a^+ = \max(a, 0)\).

The key idea of simulated annealing follows from the observation that the kernel \(p^{V, \beta}\) admits an unique equilibrium probability measure which tends to the uniform distribution on the set \(E_{\min}\) of global minima of \(V\) as \(\beta \rightarrow \infty\). On can then expect asymptotic convergence to a configuration \(z \in E_{\min}\). Indeed, early results [6-8] show that this desirable property holds for suitably adjusted logarithmic cooling schedules of the form \(\beta_n = \beta_0 \ln(n+1)\). However, in practice, one always deals with a finite amount of computing time and it has been demonstrated more recently [9] that exponential cooling schedules \(\beta_n = \beta_0 \exp(n \xi), 1 \leq n \leq N\), are to be preferred (as verified experimentally in [3, 4]).

3 Image restoration in a wavelet basis

The finite time annealing algorithms \((\Delta^S, U, [g]_{\Delta^S}, \nu_0, (\beta_n)_{n \leq N})\), where \([g]_{\Delta^S}\) is given by (4.5), are generally reported to perform poorly. As shown in [2-4], significant improvements can be obtained by restricting the digital image space to a locally bounded image subspace \(\Omega^S(\delta)\) which consists of all configurations \(x \in \Delta^S\) such that, \(\forall s \in \mathcal{S}\),

\[
\min_{r \in \mathcal{N}^E_s} x_r - \delta \leq x_s \leq \max_{r \in \mathcal{N}^E_s} x_r + \delta. \tag{8}
\]

where \(\Delta^S\) is the 4-nearest neighbor system on \(\mathcal{S}\). Still, due to the fact that the sites are updated singly, it remains extremely difficult to move between high probability areas of the state space separated by low probability regions. Clearly, one can find a solution for this problem by switching the values of a whole block of pixels. Since this turns out to be very impractical in the spatial domain, we propose to carry out this operation through single-site updating of the DWT coefficients.

In a wavelet basis, an image \(x \in \Delta^S\) is represented by \(\{x_j, w_j^i; j = 1, \ldots, J, i = 1, 2, 3\}\), where \(x_j\) is the approximation of \(x\) at resolution \(2^{-j}\) and the difference of information between \(x_{j-1}\) and the coarser approximation \(x_j\) is given by the detail images \(\{w_j^i; i = 1, 2, 3\}\). By arranging the subimages \(\{x_j, w_j^i\}\) in a column vector \(\hat{x}\), the basis transfer scheme can be simply implemented with an analysis and synthesis matrix pair [10]:

\[
\hat{x} = \mathcal{A}x, \quad x = \mathcal{S}\hat{x}, \quad \hat{x} = [x_j, w_j^1, w_j^2, w_j^3, \ldots, w_j^1, w_j^2, w_j^3]^t. \tag{9}
\]

Hence it follows that the estimate of the true image also writes

\[\hat{x}^0 = \mathcal{S} \arg \min_{x \in \mathcal{X}} \hat{U}(\hat{x}), \quad \hat{U}(\hat{x}) = U(\mathcal{S}\hat{x}). \tag{10}\]

Without loss of generality, we will assume that the size of \(\mathcal{S}\) is \(M \times M\), where \(M/2^J = [M/2^J] > 1\). If \(x^0\) is piecewise smooth, it is fair to restrict the DWT domain to the set

\[
\tilde{\Delta}^S = \Delta^S/4, \quad \tilde{\mathcal{A}} = \begin{pmatrix} J \times 3 \times \Lambda_{i,j} / 4 \end{pmatrix}, \quad \Lambda_{i,j} = [-|w_j^i|_{\max}, |w_j^i|_{\max}], \tag{11}
\]

where the maximum absolute value \(|w_j^i|_{\max}\) of the coefficients of a particular subband can be computed from the maximum amplitude of the lowest order discontinuities that are to be found in \(x^0\).
Since we deal with finite state spaces, the wavelet coefficients have to be quantized; their distributions being a priori unknown, we use a linear scalar uniform quantiser. Let $N_j^i$ be the number of quantization bins associated with $w_j^i$, $N_J$ corresponding to the low resolution residual. The resulting DWT subspace is then made up of all configurations $\tilde{x} \in \Lambda^S$ such that

$$
\tilde{x} = \left\{ \lambda_{\min} + \frac{\lambda_{\max} - \lambda_{\min}}{N_j^i} \left( \lambda_j + \frac{1}{2} \right),  \\
|w_j^i|_{\max} \left( \frac{2w_j^i + 1}{N_j^i} - 1 \right) \right\},
$$

where $\tilde{\Lambda} = \{ \tilde{\Lambda}_j, \tilde{\Lambda}_j^i \}$ belongs to a set $\tilde{\Lambda}^S$ similar to $\Lambda^S$ (11) with $\Lambda$ and $\Lambda_j^i$ respectively replaced by $\tilde{\Lambda} = \{0, \ldots, N_j - 1\}$ and $\Lambda_j^i = \{0, \ldots, N_j^i - 1\}$. Now, image restoration can be performed in a wavelet basis by means of the new class of algorithms $(\tilde{\Lambda}^S, \tilde{U}, [q]_{\tilde{\Lambda}^S}, \nu_0, (\beta_n^N)_{1 \leq n \leq N})$, where $[q]_{\tilde{\Lambda}^S}$ is given by (4) when $E = \tilde{\Lambda}^S$.

In order to encourage better sampling of the state space, an immediate additional refinement consists in restricting $\Lambda^S$ to some subset $\Omega^S$ of $(\delta_j)$. Besides, recall that the detail subbands $w_j^i$ can be interpreted as spatially oriented high frequency channels in a vertical ($i = 1$), horizontal ($i = 2$) or diagonal ($i = 3$) direction. Keeping this in mind, let $N_1$ and $N_2$ be respectively the 2-nearest horizontal and vertical neighbor systems, and let $N_3$ be the 4-nearest diagonal neighbor system. It clearly comes up that a locally bounded image subspace defined by $\Lambda^S$ leads to a faithful description of the configurations in $\Lambda_j^i$. The restricted DWT space considered for minimizing $\tilde{U}$ then writes

$$
\tilde{\Omega}^S = \Omega^S \times \left\{ j \times \left( \frac{3}{j} \right) \times \Omega_j^S \times (\delta_j) \right\}
$$

and it is straightforward to construct an irreducible symmetric Markov kernel $[q]_{\tilde{\Omega}^S}$ on $\tilde{\Omega}^S$ such that the annealing theoretical convergence results apply to the class of algorithms

$$
(\tilde{\Omega}^S, \tilde{U}, [q]_{\tilde{\Omega}^S}, \nu_0, (\beta_n^N)_{1 \leq n \leq N}).
$$

The proposed minimization scheme is particularly well suited for images that are dominated by their low frequency portion. As an example, consider, the $72 \times 72$ piecewise planar original image depicted in fig. 1(a). The data (fig. 1(b)) were generated with a shift-variant Gaussian blur and a white Gaussian noise specified by a 18 dB blurred signal-to-noise ratio (BSNR); the blur varies linearly in the radial direction, starting at an initial standard deviation of 1.0 in the center of the image, and ending at a standard deviation of 5.0 at the image corner. Figures 2(a) and 2(b) display some restoration results obtained by minimizing (2) $(k = 2, \phi(a) = |a|/(1 + |a|))$ with a Metropolis-type anneal algorithms of the form

$$
(\Omega^S(\delta), U, [q]_{\Omega^S(\delta)}, \nu_0, (\beta_0 \exp(n|\xi|))_{1 \leq n \leq N})
$$

working on the subspace $\Omega^S(5)$ defined over 128 gray levels. The restored image in fig. 2(a) corresponds to the choice $(\lambda_2, \Delta_2) = (8.0, 1.0)$ whereas $(\lambda_2, \Delta_2) = (16.0, 10.0)$ in the case of fig. 2(b). Though one may find that setting $\Delta_2 = 1.0$ is not very clever for this particular example, the sensitivity to the choice of the hyperparameters is plainly demonstrated. Concurrently, the estimates shown in figures 2(c) and 2(d) result from the minimization of the same cost functionals in a biorthogonal wavelet basis [11] at a resolution level $J = 2$, using an algorithm of type (14) and an exponential cooling schedule. We considered a subspace $\Omega^S$ defined by $N_J = 128$, $\delta_J = 5$, $N_j^0 \leq 63$ and $\delta_j^i \leq 5$, the exact value of the latter two depending on $|w_j^i|_{\max}$. On top of leading to much higher quality results, the proposed approach shows less sensitivity to the choice of the hyperparameters. Finally, the convergence rates achieved with the two algorithms appear in figure 3 for the case in which $(\lambda_2, \Delta_2) = (8.0, 1.0)$. Note that both algorithms were randomly initialized and that the cooling schedules were selected according to the methods set out in [4]. It appears that (15) gets easily stucked in some poor local minima while (14) allows to reach much lower energy levels.
4 “Implicitly interacting” multiresolution discontinuities

Here, we briefly discuss a somewhat promising way for taking into account the mutual dependence between neighboring horizontal or vertical discontinuities. In particular, our approach incorporates the smoothness features of the line field implicitly rather than explicitly (see, e.g., [6]).

Due to the spatial orientations of the high frequency subbands, the prolongation of discontinuities at different resolution levels can be facilitated in a natural way through the introduction of a penalty term

\[
\Psi(\tilde{x}) = \sum_{J=1}^{L_j} \sum_{i=1}^{S_j} \tilde{D}_j^i \left( [\tilde{D}_j^i w_j^i] / \Delta_j \right),
\]

where \(\tilde{D}_j^i \in \mathbb{R}^{L_j \times S_j / 4^j}\) is a first-order discrete derivative operator in an horizontal \((i = 1)\) or vertical \((i = 2)\) direction. Then, one is left with the choice of minimizing either \(U(\tilde{x}) + \lambda \Psi(\tilde{x})\) using (14) or \(U(x) + \lambda \tilde{\Psi}(x)\) using (15). Though the introduction of additional hyper-parameters is a bit awkward, this approach does not require other heuristic specifications such as the costs associated to different edge configurations [6]. Furthermore, besides from being suitable to any order of discontinuity, it is truly multiresolution and it provides an interesting alternative to the methods devised in [12] and [13].

The potential of the proposed scheme is clearly illustrated in fig. 4. The original piecewise constant image (fig. 4(a)) was degraded with the same shift-variant Gaussian kernel as for the previous example and a 12 dB BSNR white Gaussian noise (fig. 4(b)). Figure 4(c) displays the estimate resulting from the minimization of \(U(2)(3) (k = 1, \phi(a) = |a|/(1 + |a|))\) using a simulated annealing algorithm defined on a restricted spatial image space (15). The restoration result shown in fig. 4(d) was obtained in a similar way by minimizing the augmented functional \(U(x) + \lambda \tilde{\Psi}(x)\) with \(\tilde{\Psi}\) operating in the spatial frequency domain at a resolution level \(J = 2\) and \(\phi(a) = |a|\).

5 Conclusion

We have considered the shift-variant image restoration problem using Metropolis-type annealing algorithms; it was demonstrated that two main drawbacks of the usual approach can be overcome in a multiresolution framework without altering the theoretical convergence properties of annealing. In the case of piecewise smooth images with second or higher order
discontinuities, it turns out that convergence towards global minima of the energy landscape is highly improved when minimization is performed on a restricted DWT domain. One should notice here that further improvements can be achieved through increasing concave transform of the cost functional [3, 4]. We also introduced the idea of implicitly interacting multiresolution discontinuities so that the prolongation of edges at different resolution levels takes place in a quite natural way.

References


