

Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising

I. Daubechies* and G. Teschke†

December 2, 2004

Abstract

Inspired by papers of Vese–Osher [20] and Osher–Solé–Vese [19] we present a wavelet–based treatment of variational problems arising in the field of image processing. In particular, we follow their approach and discuss a special class of variational functionals that induce a decomposition of images into oscillating and cartoon components and possibly an appropriate ‘noise’ component. In the setting of [20] and [19], the cartoon component of an image is modeled by a BV function; the corresponding incorporation of BV penalty terms in the variational functional leads to PDE schemes that are numerically intensive. By replacing the BV penalty term by a $B_1^1(L_1)$ term (which amounts to a slightly stronger constraint on the minimizer), and writing the problem in a wavelet framework, we obtain elegant and numerically efficient schemes with results very similar to those obtained in [20] and [19]. This approach allows us, moreover, to incorporate general bounded linear blur operators into the problem so that the minimization leads to a simultaneous decomposition, deblurring and denoising.

1 Introduction

One important problem in image processing is the restoration of the ‘true’ image from an observation. In almost all applications the observed image is a noisy and blurred version of the true image. In principle, the restoration task can be understood as an inverse problem, i.e. one can attack it by solving a related variational problem.

In this paper we focus on a special class of variational problems which induce a decomposition of images in oscillating and cartoon components; the cartoon part is ideally piecewise smooth with possible abrupt edges and contours; the oscillation part on the other hand ‘fills’ in the smooth regions in the cartoon with texture -like features. Several authors, e.g. [20, 19], propose to model the cartoon component by the space BV which induces a penalty term that allows edges and contours in the

*Princeton University, PACM, Washington Road, Princeton, NJ 08544-1000, USA

†University of Bremen, ZETEM, P.O.Box 330 440, 28334 Bremen, Germany

reconstructed cartoon images. However, the minimization of variational problems of this type usually results in PDE based schemes that are numerically intensive.

The main goal of this paper is to provide a computationally thriffter algorithm by using a wavelet-based scheme that solves not the same but a very similar variational problem, in which the BV -constraint, which cannot easily be expressed in the wavelet domain, is replaced by a $B_1^1(L_1)$ -term, i.e. a slightly stricter constraint (since $B_1^1(L_1) \subset BV$ in two dimensions). Moreover, we can allow the involvement of general linear bounded blur operators, which extends the range of application. By applying recent results, see [7], we show convergence of the proposed scheme.

In order to give a brief description of the underlying variational problems, we recall the methods proposed in [20, 19]. They follow the idea of Y. Meyer [18], proposed as an improvement on the total variation framework of L. Rudin, S. Osher and E. Fatemi [21]. In principle, the models can be understood as a decomposition of an image f into $f = u + v$, where u represents the cartoon part and v the texture part. In the Vese-Osher model, see [20], the decomposition is induced by solving

$$\inf_{u, g_1, g_2} G_p(u, g_1, g_2) , \quad \text{where} \quad (1.1)$$

$$G_p(u, g_1, g_2) = \int_{\Omega} |\nabla u| + \lambda \|f - (u + \operatorname{div} g)\|_{L_2(\Omega)}^2 + \mu \| \|g\| \|_{L_p(\Omega)} ,$$

with $f \in L_2(\Omega)$, $\Omega \subset \mathbb{R}^2$, and $v = \operatorname{div} g = \operatorname{div}(g_1, g_2)$. The first term is the total variation of u . If $u \in L_1$ and $|\nabla u|$ is a finite measure on Ω , then $u \in BV(\Omega)$. This space allows discontinuities, therefore edges and contours generally appear in u . The second term represents the restoration discrepancy; to penalize v , the third term approximates (by taking p finite) the norm of the space of oscillating functions introduced by Y. Meyer (with $p = \infty$) which is in some sense dual to $BV(\Omega)$. (For details we refer the reader to [18].) Setting $p = 2$ and $g = \nabla P + Q$, where P is a single-valued function and Q is a divergence-free vector field, it is shown in [19] that the v -penalty term can be expressed by

$$\| \|g\| \|_{L_2(\Omega)} = \left(\int_{\Omega} |\nabla(\Delta)^{-1} v|^2 \right)^{1/2} = \|v\|_{H^{-1}(\Omega)} .$$

(The H^{-1} calculus is allowed as long as we deal with oscillatory texture/noise components that have zero mean.) With these assumptions, the variational problem (1.1) simplifies to solving

$$\inf_{u, g_1, g_2} G_2(u, v) , \quad \text{where} \quad (1.2)$$

$$G_2(u, v) = \int_{\Omega} |\nabla u| + \lambda \|f - (u + v)\|_{L_2(\Omega)}^2 + \mu \|v\|_{H^{-1}(\Omega)} .$$

In general, one drawback is that the minimization of (1.1) or (1.2) leads to numerically intensive schemes.

Instead of solving problem (1.2) by means of finite difference schemes, we propose a wavelet-based treatment. We are encouraged by the fact that elementary methods based on wavelet shrinkage solve similar extremal problems where $BV(\Omega)$ is replaced by the Besov space $B_1^1(L_1(\Omega))$. Since $BV(\Omega)$ can not be simply described in terms

of wavelet coefficients, it is not clear that $BV(\Omega)$ minimizers can be obtained in this way. Yet, it is shown in [2], exploiting $B_1^1(L_1(\Omega)) \subset BV(\Omega) \subset B_1^1(L_1(\Omega))$ – weak, that methods using Haar systems provide near $BV(\Omega)$ minimizers. So far there exists no similar result for general (in particular smoother) wavelet systems. We shall nevertheless use wavelets that have more smoothness/vanishing moments than Haar wavelets, because we expect them to be better suited to the modeling of the smooth parts in the cartoon image. Though we may not obtain provable ‘near–best– BV –minimizers’, we hope to nevertheless not be ‘too far off’. Limiting ourselves to the case $p = 2$, replacing $BV(\Omega)$ by $B_1^1(L_1(\Omega))$, and, moreover, extending the range of applicability by incorporating a bounded linear operator K , we end up with the following variational problem:

$$\inf_{u,v} \mathcal{F}_f(v, u) , \quad \text{where}$$

$$\mathcal{F}_f(v, u) = \|f - K(u + v)\|_{L_2(\Omega)}^2 + \gamma \|v\|_{H^{-1}(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))} .$$

This paper is organized as follows. In Section 2 we recall some basic facts on wavelets, in Section 3 the numerical scheme is developed and convergence is shown, in Section 4 we introduce some extra refinements on the scheme, and finally, in Section 5 we present some numerical results.

2 Preliminaries on wavelets

In this section, we briefly recall some facts on wavelets that are needed later on. Especially important for our approach are the smoothness characterization properties of wavelets: one can determine the membership of a function in many different smoothness functional spaces by examining the decay properties of its wavelets coefficients. For a comprehensive introduction and overview on this topic we would refer the reader to the abundant literature, see e.g. [5, 6, 1, 4, 13, 12, 15, 23].

Suppose H is a Hilbert space. Let $\{V_j\}$ be a sequence of closed nested subspaces of H whose union is dense in H while their intersection is zero. In addition, V_0 is shift-invariant and $f \in V_j \leftrightarrow f(2^j \cdot) \in V_0$, so that the sequence $\{V_j\}$ forms a multi-resolution analysis. In many cases of practical relevance the spaces V_j are spanned by single scale bases $\Phi_j = \{\phi_{j,k} : k \in I_j\}$ which are uniformly stable. Successively updating a current approximation in V_j to a better one in V_{j+1} can be facilitated if stable bases $\Psi_j = \{\psi_{j,k} : k \in J_j\}$ for some complement W_j of V_j in V_{j+1} are available. Hence, any $f_n \in V_n$ has an alternative multi-scale representation $f_n = \sum_{k \in I_0} f_{0,k} \phi_{0,k} + \sum_{j=0}^n \sum_{k \in J_j} f_{j,k} \psi_{j,k}$. The essential constraint on the choice of W_j is that $\Psi = \bigcup_j \Psi_j$ forms a Riesz-basis of H , i.e. every $f \in H$ has a unique expansion

$$f = \sum_j \sum_{k \in J_j} \langle f, \tilde{\psi}_{j,k} \rangle \psi_{j,k} \quad \text{such that} \quad \|f\|_H \sim \left(\sum_j \sum_{k \in J_j} |\langle f, \tilde{\psi}_{j,k} \rangle|^2 \right)^{\frac{1}{2}} , \quad (2.1)$$

where $\tilde{\Psi}$ forms a bi-orthogonal system and is in fact also a Riesz-basis for H , see, e.g., [5].

For our approach we assume that any function (image) $f \in L_2(I)$ can be extended periodically to all of \mathbb{R}^2 . Here I is assumed to be the unit square $(0, 1]^2 = \Omega$. Throughout this paper we only consider compactly supported tensor product wavelet systems (based on Daubechies' orthogonal wavelets, see [6], or symmetric bi-orthogonal wavelets by Cohen, Daubechies, and Feauveau, see [1]).

We are finally interested in characterizations of Besov spaces, see, e.g., [23]. For $\beta > 0$ and $0 < p, q \leq \infty$ the Besov space $B_q^\beta(L_p(\Omega))$ of order β is the set of functions

$$B_q^\beta(L_p(\Omega)) = \{f \in L_p(\Omega) : |f|_{B_q^\beta(L_p(\Omega))} < \infty\},$$

where $|f|_{B_q^\beta(L_p(\Omega))} = \left(\int_0^\infty (t^{-\beta} \omega_l(f; t)_p)^q dt/t\right)^{1/q}$ and ω_l denotes the l -th modulus of smoothness, $l > \beta$. These spaces are endowed with the norm $\|f\|_{B_q^\beta(L_p(\Omega))} = \|f\|_{L_p(\Omega)} + |f|_{B_q^\beta(L_p(\Omega))}$. (For $p < 1$, this is not a norm, strictly speaking, and the Besov spaces are complete topological vector spaces but no longer Banach spaces, see [11] for details, including the characterization of these spaces by wavelets.) What is important to us is that one can determine whether a function is in $B_q^\beta(L_p(\Omega))$ simply by examining its wavelet coefficients. The case $p = q$, on which we shall focus, is the easiest. Suppose that ϕ has R continuous derivatives and ψ has vanishing moments of order M . Then, as long as $\beta < \min(R, M)$, one has in, two dimensions, for all $f \in B_p^\beta(L_p(\Omega))$, the following norm equivalence (denoted by \sim)

$$|f|_{B_p^\beta(L_p(\Omega))} \sim \left(\sum_{\lambda} 2^{|\lambda|s p} |f_{\lambda}|^p \right)^{1/p} \quad \text{with } f_{\lambda} := \langle f, \tilde{\psi}_{\lambda} \rangle, \quad s = \beta + 1 - 2/p \text{ and } |\lambda| = j. \quad (2.2)$$

In what follows, we shall always use the equivalent weighted ℓ_p -norm of the $\{f_{\lambda}\}$ instead of the standard Besov norm; with a slight abuse of notation we shall continue to denote it by the same symbol, however. When $p = q = 2$, the space $B_2^\beta(L_2(\Omega))$ is the Bessel potential space $H^\beta(\Omega)$. In analogy with the special case of Bessel potential spaces $H^\beta(\Omega)$, the Besov space $B_p^\beta(L_p(\Omega))$ with $\beta < 0$ can be viewed as the dual space of $B_{p'}^{\beta'}(L_{p'}(\Omega))$, where $\beta' = -\beta$ and $1/p + 1/p' = 1$.

3 Image decomposition

As stated in Section 1, we aim to solve

$$\inf_{u,v} \mathcal{F}_f(v, u), \quad \text{where} \quad (3.1)$$

$$\mathcal{F}_f(v, u) = \|f - K(u + v)\|_{L_2(\Omega)}^2 + \gamma \|v\|_{H^{-1}(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))}.$$

At first, we may observe the following

Lemma 3.1 *If the null-space $\mathcal{N}(K)$ of the operator K is trivial, then the variational problem (3.1) has a unique minimizer.*

This can be seen as follows:

$$\mathcal{F}_f(\mu(v, u) + (1 - \mu)(v', u')) - \mu \mathcal{F}_f((v, u)) - (1 - \mu) \mathcal{F}_f((v', u')) =$$

$$\begin{aligned}
& -\mu(1-\mu) \left(\|K(u-u'+v-v')\|_{L_2(\Omega)}^2 + \gamma \|v-v'\|_{H^{-1}(\Omega)}^2 \right) \\
& + 2\alpha \left(|\mu u + (1-\mu)u'|_{B_1^1(L_1(\Omega))} - \mu|u|_{B_1^1(L_1(\Omega))} - (1-\mu)|u'|_{B_1^1(L_1(\Omega))} \right) \quad (3.2)
\end{aligned}$$

with $0 < \mu < 1$. Since the Banach norm is convex the right hand side of (3.2) is non-positive, i.e. \mathcal{F}_f is convex. Since $\mathcal{N}(K) = \{0\}$, the term $\|K(u-u'+v-v')\|$ can be zero only if $u-u'+v-v'=0$, moreover, $\|v-v'\|$ is zero only if $v-v'=0$. Hence, (3.2) is strictly convex. \blacksquare

In order to solve this problem by means of wavelets we have to switch to the sequence space formulation. When K is the identity operator the problem simplifies to

$$\inf_{u,v} \left\{ \sum_{\lambda \in J} (|f_\lambda - (u_\lambda + v_\lambda)|^2 + \gamma 2^{-2|\lambda|} |v_\lambda|^2 + 2\alpha |u_\lambda|) \right\}, \quad (3.3)$$

where $J = \{\lambda = (i, j, k) : k \in J_j, j \in \mathbb{Z}, i = 1, 2, 3\}$ is the index set used in our separable setting. The minimization of (3.3) is straightforward, since it decouples into easy one-dimensional minimizations. This results in an explicit shrinkage scheme, presented also in [8]:

Proposition 3.1 *Let f be a given function. The functional (3.3) is minimized by the parametrized class of functions $\tilde{v}_{\gamma,\alpha}$ and $\tilde{u}_{\gamma,\alpha}$ given by the following non-linear filtering of the wavelet series of f :*

$$\tilde{v}_{\gamma,\alpha} = \sum_{\lambda \in J_{j_0}} (1 + \gamma 2^{-2|\lambda|})^{-1} [f_\lambda - S_{\alpha(2^{2|\lambda|} + \gamma)/\gamma}(f_\lambda)] \psi_\lambda$$

and

$$\tilde{u}_{\gamma,\alpha} = \sum_{k \in I_{j_0}} \langle f, \tilde{\phi}_{j_0,k} \rangle \phi_{j_0,k} + \sum_{\lambda \in J_{j_0}} S_{\alpha(2^{2|\lambda|} + \gamma)/\gamma}(f_\lambda) \psi_\lambda,$$

where S_t denotes the soft-shrinkage operator, J_{j_0} all indices λ for scales larger than j_0 and I_{j_0} the indices λ for the fixed scale j_0 .

In the case where K is not the identity operator the minimization process results in a coupled system of nonlinear equations for the wavelet coefficients u_λ and v_λ , which is not as straightforward to solve. To overcome this problem, we adapt an iterative approach. As in [7] we derive the iterative algorithm from a sequence of so-called surrogate functionals that are each easy to minimize, and for which one hopes that the successive minimizers have the minimizing element of (3.1) as limit. However, contrary to [7] our variational problem has mixed quadratic and non-quadratic penalties. This requires a slightly different use of surrogate functionals. In [10, 9] a similar $u+v$ problem is solved by an approach that combines u and v into one vector-valued function (u, v) . This leads to alternating iterations with respect to u and v simultaneously. It can be shown that the minimizers of the resulting alternating algorithm strongly converge to the desired unique solution, [10].

We will follow a different approach here, in which we first solve the quadratic problem for v , and then construct an iteration scheme for u . To this end, we

introduce the differential operator $T := (-\Delta)^{1/2}$. Setting $v = Th$ the variational problem (3.1) reads as

$$\inf_{(u,h)} \mathcal{F}_f(h, u) , \quad \text{with} \quad (3.4)$$

$$\mathcal{F}_f(h, u) = \|f - K(u + Th)\|_{L_2(\Omega)}^2 + \gamma \|h\|_{L_2(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))} .$$

Minimizing (3.4) with respect to w results in

$$\tilde{h}_\gamma(f, u) = (T^* K^* K T + \gamma)^{-1} T^* K^* (f - K u)$$

or equivalently

$$\tilde{v}_\gamma(f, u) = T(T^* K^* K T + \gamma)^{-1} T^* K^* (f - K u) .$$

Inserting this explicit expression for $\tilde{h}_\gamma(f, u)$ in (3.4) and defining

$$f_\gamma := T_\gamma f, \quad T_\gamma^2 := I - K T (T^* K^* K T + \gamma)^{-1} T^* K^* , \quad (3.5)$$

we obtain

$$\mathcal{F}_f(\tilde{h}_\gamma(f, u), u) = \|f_\gamma - T_\gamma K u\|_{L_2(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))} . \quad (3.6)$$

Thus, the remaining task is to solve

$$\inf_u \mathcal{F}_f(\tilde{h}_\gamma(f, u), u) , \quad \text{where} \quad (3.7)$$

$$\mathcal{F}_f(\tilde{h}_\gamma(f, u), u) = \|f_\gamma - T_\gamma K u\|_{L_2(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))} .$$

The corresponding variational equations in the sequence space representation are

$$\forall \lambda : (K^* T_\gamma^2 K u)_\lambda - (K^* f_\gamma)_\lambda + \alpha \text{sign}(u_\lambda) = 0 .$$

This gives a coupled system of nonlinear equations for u_λ . For this reason we construct surrogate functionals that remove the influence of $K^* T_\gamma^2 K u$. First, we choose a constant C such that $\|K^* T_\gamma^2 K\| < C$. Since $\|T_\gamma\| \leq 1$, it suffices to require that $\|K^* K\| < C$. Then we define the functional

$$\Phi(u; a) := C \|u - a\|_{L_2(\Omega)}^2 - \|T_\gamma K(u - a)\|_{L_2(\Omega)}^2$$

which depends on an auxiliary element $a \in L_2(\Omega)$. We observe that $\Phi(u, a)$ is strictly convex in u for any a . Since K can be rescaled, we limit our analysis without loss of generality to the case $C = 1$. We finally add $\Phi(u; a)$ to $\mathcal{F}_f(\tilde{h}_\gamma(f, u), u)$ and obtain the following surrogate functional

$$\begin{aligned} \mathcal{F}_f^{sur}(\tilde{h}_\gamma(f, a), u; a) &= \mathcal{F}_f(\tilde{h}_\gamma(f, u), u) + \Phi(u; a) \\ &= \sum_\lambda \{u_\lambda^2 - 2u_\lambda(a + K^* T_\gamma^2(f - Ka))_\lambda + 2\alpha |u_\lambda|\} \\ &\quad + \|f_\gamma\|_{L_2(\Omega)}^2 + \|a\|_{L_2(\Omega)}^2 - \|T_\gamma K a\|_{L_2(\Omega)}^2 . \end{aligned} \quad (3.8)$$

The advantage of minimizing (3.8) is that the variational equations for u_λ decouple. The summands of (3.8) are differentiable in u_λ except at the point of non-differentiability. The variational equations for each λ are now given by

$$u_\lambda + \alpha \text{sign}(u_\lambda) = (a + K^* T_\gamma^2(f - Ka))_\lambda .$$

This results in an explicit soft-shrinkage operation for u_λ

$$u_\lambda = S_\alpha((a + K^*T_\gamma^2(f - Ka))_\lambda) .$$

The next proposition summarizes our findings; it is the specialization to our particular case of a more general theorem in [7].

Proposition 3.2 *Suppose K is a linear bounded operator modeling the blur, with K maps $L_2(\Omega)$ to $L_2(\Omega)$ and $\|K^*K\| < 1$. Moreover, assume T_γ is defined as in (3.5) and the functional $\mathcal{F}_f^{sur}(\tilde{h}, u; a)$ is given by*

$$\mathcal{F}_f^{sur}(\tilde{h}_\gamma(f, u), u; a) = \mathcal{F}_f(\tilde{h}_\gamma(f, u), u) + \Phi(u; a) .$$

Then, for arbitrarily chosen $a \in L_2(\Omega)$, the functional $\mathcal{F}_f^{sur}(\tilde{h}_\gamma(f, u), u; a)$ has a unique minimizer in $L_2(\Omega)$. The minimizing element is given by

$$\tilde{u}_{\gamma, \alpha} = \mathbf{S}_\alpha(a + K^*T_\gamma^2(f - Ka)) ,$$

where the operator \mathbf{S}_α is defined component-wise by

$$\mathbf{S}_\alpha(x) = \sum_\lambda S_\alpha(x_\lambda)\psi_\lambda .$$

The proof follows from [7]. One can now define an iterative algorithm by repeated minimization of \mathcal{F}_f^{sur} :

$$u^0 \text{ arbitrary ; } u^n = \arg \min_u \left(\mathcal{F}_f^{sur}(\tilde{h}_\gamma(f, u), u; u^{n-1}) \right) \quad n = 1, 2, \dots \quad (3.9)$$

The convergence result of [7] can again be applied directly:

Theorem 3.1 *Suppose K is a linear bounded operator, with $\|K^*K\| < 1$, and that T_γ is defined as in (3.5). Then the sequence of iterates*

$$u_{\gamma, \alpha}^n = \mathbf{S}_\alpha(u_{\gamma, \alpha}^{n-1} + K^*T_\gamma^2(f - Ku_{\gamma, \alpha}^{n-1})) , \quad n = 1, 2, \dots ,$$

with arbitrarily chosen $u^0 \in L_2(\Omega)$, converges in norm to a minimizer $\tilde{u}_{\gamma, \alpha}$ of the functional

$$\mathcal{F}_f(\tilde{h}_\gamma(f, u), u) = \|T_\gamma(f - Ku)\|_{L_2(\Omega)}^2 + 2\alpha|u|_{B_1^1(L_1(\Omega))} .$$

If $\mathcal{N}(T_\gamma K) = \{0\}$, then the minimizer $\tilde{u}_{\gamma, \alpha}$ is unique, and every sequence of iterates converges to $\tilde{u}_{\gamma, \alpha}$ in norm.

Combining the result of Theorem 3.1 and the representation for \tilde{v} we summarize how the image can finally be decomposed in cartoon and oscillating components.

Corollary 3.1 *Assume that K is a linear bounded operator modeling the blur, with $\|K^*K\| < 1$. Moreover, if T_γ is defined as in (3.5) and if $\tilde{u}_{\gamma, \alpha}$ is the minimizing element of (3.7), obtained as a limit of $u_{\gamma, \alpha}^n$ (see Theorem 3.1), then the variational problem*

$$\inf_{(u, h)} \mathcal{F}_f(h, u), \text{ with } \mathcal{F}_f(h, u) = \|f - K(u + Th)\|_{L_2(\Omega)}^2 + \gamma\|h\|_{L_2(\Omega)}^2 + 2\alpha|u|_{B_1^1(L_1(\Omega))}$$

is minimized by the class

$$(\tilde{u}_{\gamma, \alpha}, (T^*K^*KT + \gamma)^{-1}T^*K^*(f - K\tilde{u}_{\gamma, \alpha})) .$$

where $\tilde{u}_{\gamma, \alpha}$ is the unique limit of the sequence

$$u_{\gamma, \alpha}^n = \mathbf{S}_\alpha(u_{\gamma, \alpha}^{n-1} + K^*T_\gamma^2(f - Ku_{\gamma, \alpha}^{n-1})) , \quad n = 1, 2, \dots .$$

4 Refinements: using redundancy and adaptivity to reduce artifacts

The non-linear filtering rule of Proposition 3.1 gives explicit descriptions of \tilde{v} and \tilde{u} that are computed by fast discrete wavelet schemes. However, non-redundant filtering very often creates artifacts in terms of undesirable oscillations, which manifest themselves as ringing and edge blurring. Poor directional selectivity of traditional tensor product wavelet bases likewise cause artifacts. In this section we discuss various refinements on the basic algorithm that address this problem. In particular, we shall use redundant translation invariant schemes, complex wavelets, and additional edge dependent penalty weights.

4.1 Translation invariance by cycle-spinning

Assume that we are given an image with 2^M rows of 2^M pixels, where the gray value of each pixel gives an average of f on a square $2^{-M} \times 2^{-M}$, which we denote by f_k^M , with k a double index running through all the elements of $\{0, 1, \dots, 2^M - 1\} \times \{0, 1, \dots, 2^M - 1\}$. A traditional wavelet transform then computes f_l^j , $d_l^{j,i}$ with $j_0 \leq j \leq M$, $i = 1, 2, 3$ and $l \in \{0, 1, \dots, 2^j - 1\} \times \{0, 1, \dots, 2^j - 1\}$ for each j , where the f_l^j stand for an average of f on mostly localized on (and indexed by) the squares $[l_1 2^{-j}, (l_1 + 1) 2^{-j}] \times [l_2 2^{-j}, (l_2 + 1) 2^{-j}]$, and the $d_l^{j,i}$ stand for the different species of wavelets (in two dimensions, there are three) in the tensor product multi-resolution analysis. Because the corresponding wavelet basis is not translation invariant, Coifman and Donoho proposed in [3] to recover translation invariance by averaging over the $2^{2(M+1-j_0)}$ translates of the wavelet basis; since many wavelets occur in more than one of these translated bases (in fact, each $\psi_{j,i,k}(x - 2^M n)$ in exactly $2^{2(j+1-j_0)}$ different bases), the average over all these bases uses only $(M + 1 - j_0) 2^{2M}$ different basis functions (and not $2^{4(M+1-j_0)}$ = number of bases \times number of elements in each basis). This approach is called *cycle-spinning*. Writing, with a slight abuse of notation, $\psi_{j,i,k+2^{j-M}n}$ for the translate $\psi_{j,i,k}(x - 2^M n)$, this average can then be written as

$$f^M = 2^{-2(M+1-j_0)} \sum_{l_1, l_2=0}^{2^M-1} \left\{ f_{l_2 2^{-M+j_0}}^{j_0} \phi_{j_0, l_2 2^{-M+j_0}} + \sum_{j=j_0}^{M-1} 2^{2(j-j_0)} \sum_{i=1}^3 d_{l_2 2^{-M+j}}^{j,i} \psi_{j,i, l_2 2^{-M+j}} \right\}.$$

Carrying out our nonlinear filtering in each of the bases and averaging the result then corresponds to applying the corresponding nonlinear filtering on the (much smaller number of) coefficients in the last expression. This is the standard way to implement thresholding on cycle-spinned representations.

The resulting sequence space representation of the variational functional (3.3) has to be adapted to the redundant representation of f . To this end, we note that the Besov penalty term takes the form

$$|f|_{B_p^\beta(L_p)} \sim \left(\sum_{j \geq j_0, i, k} 2^{(js+2(j-M))} |\langle f, \tilde{\psi}_{j,i,k} 2^{j-M} \rangle|^p \right)^{1/p}.$$

The norms $\|\cdot\|_{L_2}^2$ and $\|\cdot\|_{H^{-1}}^2$ change similarly. Consequently, we obtain the same minimization rule but with respect to a richer class of wavelet coefficients.

4.2 Directional sensitivity by frequency projections

It has been shown by several authors [16, 22, 14] that if one treats positive and negative frequencies separately in the one-dimensional wavelet transform (resulting in complex wavelets), the directional selectivity of the corresponding two-dimensional multi-resolution analysis is improved. This can be done by applying the following orthogonal projections:

$$\begin{aligned}\mathcal{P}^+ & : L_2 \rightarrow L_{2,+} = \{f \in L_2 : \text{supp } \hat{f} \subseteq [0, \infty)\} \\ \mathcal{P}^- & : L_2 \rightarrow L_{2,-} = \{f \in L_2 : \text{supp } \hat{f} \subseteq (-\infty, 0]\} .\end{aligned}$$

The projectors \mathcal{P}^+ and \mathcal{P}^- may be either applied to f or to $\{\phi, \tilde{\phi}\}$ and $\{\psi, \tilde{\psi}\}$. In a discrete framework these projections have to be approximated. This has been done in different ways in the literature. In [16, 22] Hilbert transform pairs of wavelets are used. In [14] f is projected (approximately) by multiplying with shifted generator symbols in the frequency domain. We follow the second approach, i.e.

$$(P^+ f)^\wedge(\omega) := \hat{f}(\omega)H(\omega - \pi/2) \quad \text{and} \quad (P^- f)^\wedge(\omega) := \hat{f}(\omega)H(\omega + \pi/2) ,$$

where f denotes the function to be analyzed and H is the low-pass filter for a conjugate quadrature mirror filter pair. One then has

$$\hat{f}(\omega) = (B^+ P^+ f)^\wedge(\omega) + (B^- P^- f)^\wedge(\omega) , \quad (4.1)$$

where the back-projections are given by

$$(B^+ f)^\wedge = \overline{\hat{f}H(\cdot - \pi/2)} \quad \text{and} \quad (B^- f)^\wedge = \overline{\hat{f}H(\cdot + \pi/2)}$$

respectively. This technique provides us with a simple multiplication scheme in Fourier, or equivalently, a convolution scheme in time domain. In a separable two-dimensional framework the projections need to be carried out in each of the two frequency variables, resulting in four approximate projection operators $P^{++}, P^{+-}, P^{-+}, P^{--}$. Because f is real, we have

$$(P^{++} f)^\wedge(-\omega) = \overline{(P^{--} f)^\wedge(\omega)} \quad \text{and} \quad (P^{+-} f)^\wedge(-\omega) = \overline{(P^{-+} f)^\wedge(\omega)} ,$$

so that the computation of $P^{-+} f$ and $P^{--} f$ can be omitted. Consequently, the modified variational functional takes the form

$$\begin{aligned}\mathcal{F}_f(u, v) & = 2 \left(\|P^{++}(f - (u + v))\|_{L_2}^2 + \|P^{+-}(f - (u + v))\|_{L_2}^2 \right) + \\ & \quad 2\lambda \left(\|P^{++}v\|_{H^{-1}}^2 + \|P^{+-}v\|_{H^{-1}}^2 \right) + 2\alpha |u|_{B_1^1(L_1)} \\ & \leq 2 \left(\|P^{++}(f - (u + v))\|_{L_2}^2 + \|P^{+-}(f - (u + v))\|_{L_2}^2 \right) + \\ & \quad 2\lambda \left(\|P^{++}v\|_{H^{-1}}^2 + \|P^{+-}v\|_{H^{-1}}^2 \right) + \\ & \quad 4\alpha \left(|P^{++}u|_{B_1^1(L_1)} + |P^{+-}u|_{B_1^1(L_1)} \right) ,\end{aligned}$$

which can be minimized with respect to $\{P^{++}v, P^{++}u\}$ and $\{P^{+-}v, P^{+-}u\}$ separately. The projections are be complex-valued, so that the thresholding operator needs to be adapted. Parameterizing the wavelet coefficients by modulus and angle and minimizing yields the following filtering rules for the projections of $\tilde{v}_{\gamma,\alpha}$ and $\tilde{u}_{\gamma,\alpha}$ (where $\cdot\cdot$ stands for any combination of $+$, $-$)

$$P^{\cdot\cdot}\tilde{v}_{\gamma,\alpha} = \sum_{\lambda \in J_{j_0}} (1 + \gamma 2^{-2|\lambda|})^{-1} [P^{\cdot\cdot}f_\lambda - S_{\alpha(2^{2|\lambda|+\gamma})/\gamma}(|P^{\cdot\cdot}f_\lambda|)e^{i\omega(P^{\cdot\cdot}f)}] \psi_\lambda$$

and

$$P^{\cdot\cdot}\tilde{u}_{\gamma,\alpha} = \sum_{k \in I_{j_0}} \langle P^{\cdot\cdot}f, \tilde{\phi}_{j_0,k} \rangle \phi_{j_0,k} + \sum_{\lambda \in J_{j_0}} (1 + \gamma 2^{-2|\lambda|})^{-1} S_{\alpha(2^{2|\lambda|+\gamma})/\gamma}(|P^{\cdot\cdot}f_\lambda|)e^{i\omega(P^{\cdot\cdot}f)} \psi_\lambda .$$

Finally, we have to apply the back-projections to obtain the minimizing functions

$$\tilde{v}_{\gamma,\alpha}^{BP} = B^{++}P^{++}\tilde{v}_{\gamma,\alpha} + B^{--}\overline{P^{++}\tilde{v}_{\gamma,\alpha}} + B^{+-}P^{+-}\tilde{v}_{\gamma,\alpha} + B^{-+}\overline{P^{+-}\tilde{v}_{\gamma,\alpha}}$$

and

$$\tilde{u}_{\gamma,\alpha}^{BP} = B^{++}P^{++}\tilde{u}_{\gamma,\alpha} + B^{--}\overline{P^{++}\tilde{u}_{\gamma,\alpha}} + B^{+-}P^{+-}\tilde{u}_{\gamma,\alpha} + B^{-+}\overline{P^{+-}\tilde{u}_{\gamma,\alpha}} .$$

4.3 Weighted penalty functions

In order to improve the capability of preserving edges we additionally introduce a positive weight sequence w_λ in the H^{-1} penalty term. Consequently, we aim at minimizing a slightly modified sequence space functional

$$\sum_{\lambda \in J} (|f_\lambda - (u_\lambda + v_\lambda)|^2 + \gamma 2^{-2|\lambda|} w_\lambda |v_\lambda|^2 + 2\alpha |u_\lambda| \cdot \mathbf{1}_{\{\lambda \in J_{j_0}\}}) . \quad (4.2)$$

The resulting texture and cartoon components take the form

$$\tilde{v}_{\gamma,\alpha}^w = \sum_{\lambda \in J_{j_0}} (1 + \gamma w_\lambda 2^{-2|\lambda|})^{-1} [f_\lambda - S_{\alpha(2^{2|\lambda|+\gamma w_\lambda})/\gamma w_\lambda}(f_\lambda)] \psi_\lambda$$

and

$$\tilde{u}_{\gamma,\alpha}^w = \sum_{k \in I_{j_0}} \langle f, \tilde{\phi}_{j_0,k} \rangle \phi_{j_0,k} + \sum_{\lambda \in J_{j_0}} S_{\alpha(2^{2|\lambda|+\gamma w_\lambda})/\gamma w_\lambda}(f_\lambda) \psi_\lambda .$$

The main goal is to introduce a control parameter that depends on the local structure of f . The local penalty weight w_λ should be large in the presence of an edge and small otherwise; the result of this weighting is to enhance the sensitivity of u near edges. In order to do this, we must first localize the edges, which we do by a procedure similar to an edge detection algorithm in [17]. This scheme rests on the analysis of the cycle-spinned wavelet coefficients f_λ at or near the same location but at different scales. We expect that the f_λ belonging to fine decomposition scales contain informations of edges (well localized) as well as oscillating components. Oscillating texture components typically show up in fine scales only; edges on the other hand leave a signature of larger wavelet coefficients through a wider range of scales. We thus apply the following not very sophisticated edge detector. Suppose

that $f \in V_M$ and j_e denotes some ‘critical’ scale, then for a certain range of scales $|\lambda| = |(i, j, k)| = j \in \{j_0, \dots, j_1 - j_e - 2, j_1 - j_e - 1\}$ we mark all positions k where $|f_\lambda|$ is larger than a level dependent threshold parameter t_j . Here the value t_j is chosen proportional to the mean value of all wavelet coefficients of level j . We say that $|f_\lambda|$ represents an edge if k was marked for all $j \in \{j_0, \dots, j_1 - j_e - 2, j_1 - j_e - 1\}$. Finally, we adaptively choose the penalty sequence by setting

$$w_\lambda = \begin{cases} \Theta_\lambda & \text{if } j \in \{M - 1, \dots, j_1 - j_e\} \text{ and } k \text{ was marked as an edge,} \\ \vartheta_\lambda & \text{otherwise,} \end{cases}$$

where ϑ_λ is close to one and Θ_λ is much larger in order to penalize the corresponding v_λ 's.

5 Numerical experiments

In this section, we present some numerical experiments obtained with our wavelet-based schemes.

We start with the case where K is the identity operator. In order to show how the nonlinear (redundant) wavelet scheme acts on piecewise constant functions we decompose a geometric image (representing cartoon components only) with sharp contours, see Figure 1. We observe that \tilde{u} represents the cartoon part very well. The texture component \tilde{v} (plus a constant for illustration purposes) contains only some very weak contour structures.

Next, we demonstrate the performance of the Haar shrinkage algorithm successively incorporating redundancy and local penalty weights. The redundancy is implemented by cycle spinning as describe in Section 4.1. The local penalty weights are computed the following way: firstly, we apply the shrinkage operator \mathbf{S} to f with a level dependent threshold (the threshold per scale is equal to two times the mean value of all the wavelet coefficients of the scale under consideration). Secondly, the non zero values of $S_{\text{threshold}}(f_\lambda)$ per scale indicate where w_λ is set to $\Theta_\lambda = 1 + C'$ (here $C' = 10$, moreover, we set w_λ equal to $\vartheta_\lambda = 1$ elsewhere). The coefficients $S_{\text{threshold}}(f_\lambda)$ for the first two scales of a segment of a woman image are visualized in Figure 2. In Figure 3, we present our numerical results. The upper row shows the original and the noisy image. The next row visualizes the results for non-redundant Haar shrinkage (Method A). The third row shows the same but incorporating cycle spinning (Method B), and the last row shows the incorporation of cycle spinning and local penalty weights. Each extension of the shrinkage method improves the results. This is also be confirmed by comparing the signal-to-noise-ratios (which is here defined as follows: $SNR(f, g) = 10 \log_{10}(\|f\|^2 / \|f - g\|^2)$), see Table 1.

The next experiment is done on a fabric image, see Figure 4. But in contrast to the examples before, we present here the use of frequency projection as introduced in Section 4.2. The numerical result shows convincingly that the texture component can be also well separated from the cartoon part.

In order to compare the performance with the Vese–Osher TV model and with the Vese–Solé–Osher H^{-1} model we apply our scheme to a woman image (the same that was used in [20, 19]), see Figure 5. We obtain very similar results as obtained with the TV model proposed in [20]. Compared with the results obtained



Figure 1: From left to right: initial geometric image f , \tilde{u} , $\tilde{v} + 150$, computed with Db3 in the translation invariant setting, $\alpha = 0.5$, $\gamma = 0.01$.



Figure 2: Left: noisy segment of a woman image, middle and right: first two scales of $\mathbf{S}(f)$ inducing the weight function w .

with the H^{-1} model proposed in [19] we observe that our reconstruction of the texture component contains much less cartoon information. In terms of computational cost we have observed that even in the case of applying cycle spinning and edge enhancement our proposed wavelet shrinkage scheme is less time consuming than the Vese–Solé–Osher H^{-1} restoration scheme, see table 2, even when the wavelet method is implemented in Matlab, which is slower than the compiled version for the Vese–Solé–Osher scheme.

We end this section with presenting an experiment where K is not the identity operator. In our particular case K is a convolution operator with Gaussian kernel. The implementation is simply done in Fourier space. The upper row in Figure 6 shows the original f and the blurred image Kf . The lower row visualizes the results: the cartoon component \tilde{u} , the texture component \tilde{v} , and the sum of both $\tilde{u} + \tilde{v}$. One may clearly see that the deblurred image $\tilde{u} + \tilde{v}$ contains (after a small number of iterations) more small scale details than Kf . This definitely shows the capabilities of the proposed iterative deblurring scheme (3.9).

Acknowledgments

The authors would like to thank L. Vese, S. Osher, C. DeMol and M. Defrise

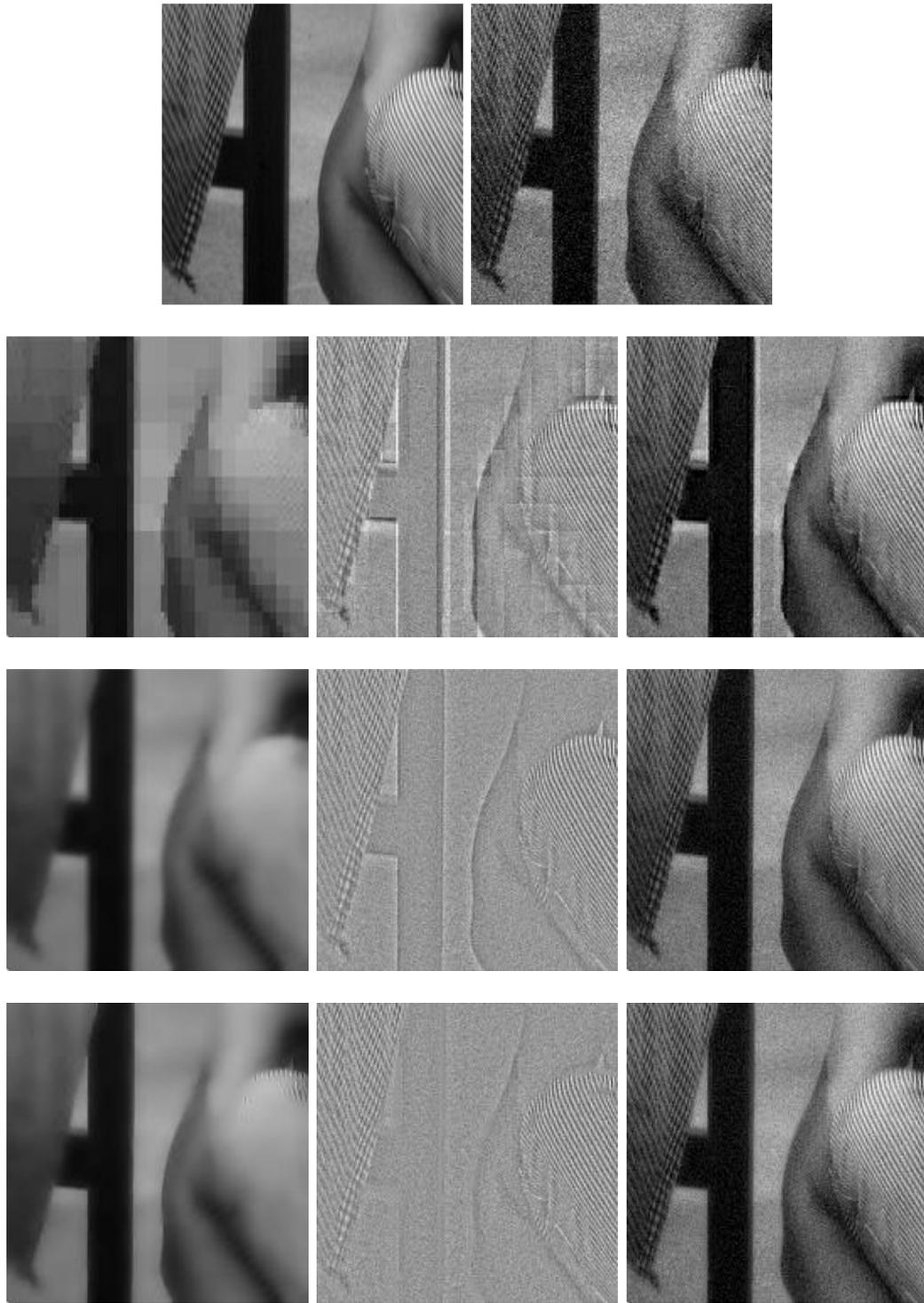


Figure 3: Top: initial and noisy image, 2nd row: non-redundant Haar shrinkage (Method A), 3rd row: translation invariant Haar shrinkage (Method B), bottom: translation invariant Haar shrinkage with edge enhancement (Method C); 2nd-4th row from left to right: \tilde{u} , $\tilde{v} + 150$ and $\tilde{u} + \tilde{v}$, $\alpha = 0.5$, $\gamma = 0.0001$, computed with Haar wavelets and critical scale $j_e = -3$.

Haar Shrinkage	$\text{SNR}(f, f_\varepsilon)$	$\text{SNR}(f, u + v)$	$\text{SNR}(f, u)$
Method A	20,7203	18,3319	16,0680
Method B	20,7203	21,6672	16,5886
Method C	20,7203	23,8334	17,5070

Table 1: Signal-to-noise ratios of the several decomposition methods (Haar shrinkage, translation invariant Haar shrinkage, translation invariant Haar shrinkage with edge enhancement).

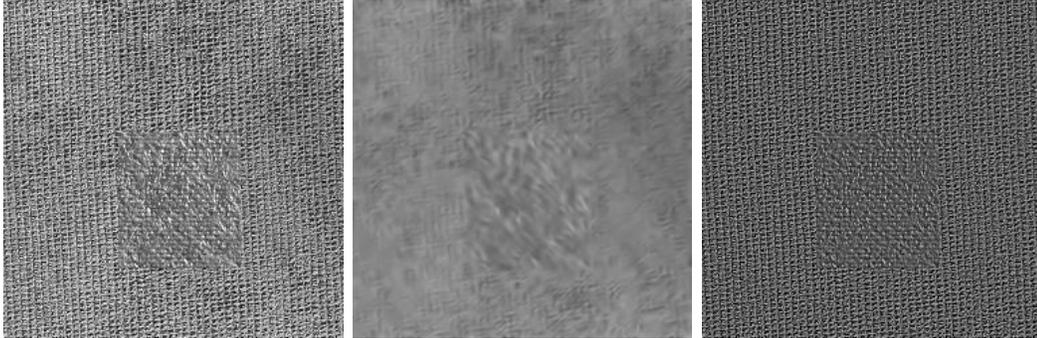


Figure 4: From left to right: initial fabric image f , \tilde{u} , $\tilde{v} + 150$, computed with Db4 incorporating frequency projections, $\alpha = 0.8$, $\gamma = 0.002$.

Data basis	"Barbara" image (512x512 pixel)
Hardware Architecture	PC
Operating System	linux
OS Distribution	redhat7.3
Model	PC, AMD Athlon-XP
Memory Size (MB)	1024
Processor Speed (MHz)	1333
Number of CPUs	1
Computational cost	(average over 10 runs)
PDE scheme in Fortran (compiler f77)	56,67 sec
wavelet shrinkage Method A (Matlab)	4,20 sec
wavelet shrinkage Method B (Matlab)	24,78 sec
wavelet shrinkage Method C (Matlab)	26,56 sec

Table 2: Comparison of computational cost of the PDE- and the wavelet-based methods.

for very interesting and stimulating discussions. We would also like to thank L. Vese for graciously lending us her program, so that we could carry out a running time comparison, as requested by one of our reviewers. I.D. gratefully acknowledges support by Air Force grant F49620-01-1-0099, and by NSF grants DMS-0219233 and DMS-0245566. G.T. gratefully acknowledges support by DAAD grants as well as by DFG grant Te 354/1-2, and moreover, he especially thanks the PACM and the

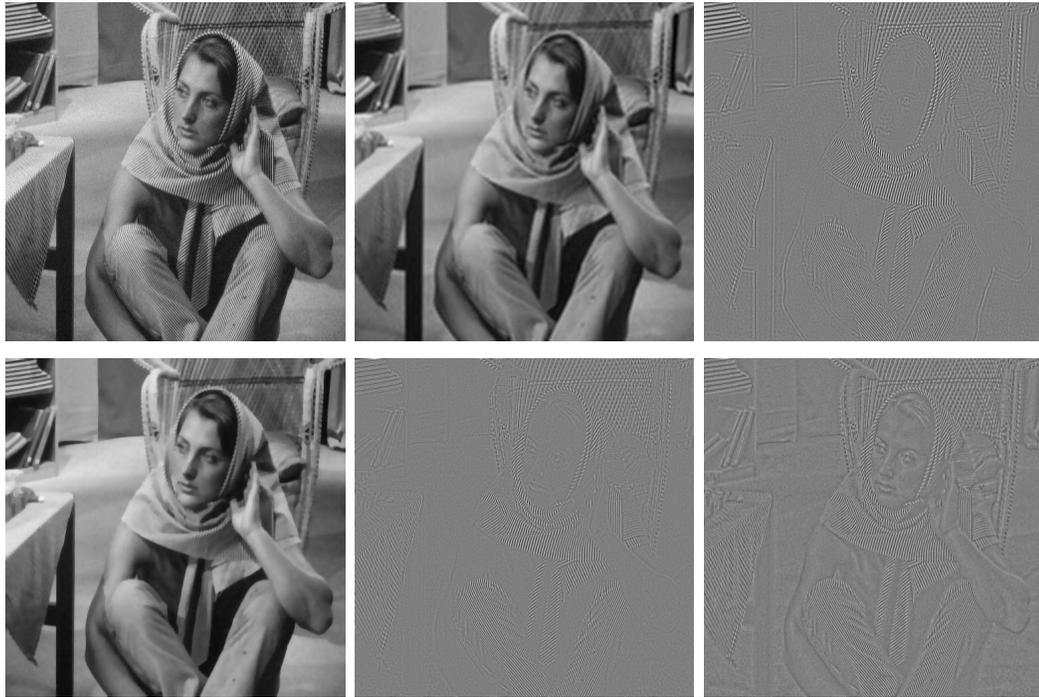


Figure 5: Top from left to right: initial woman image f , \tilde{u} and $\tilde{v} + 150$, computed with Db10 (Method C), $\alpha = 0.5$, $\gamma = 0.002$; bottom from left to right: u and v obtained by the Vese–Osher TV model and the v component obtained by the Vese–Solé–Osher H^{-1} model.

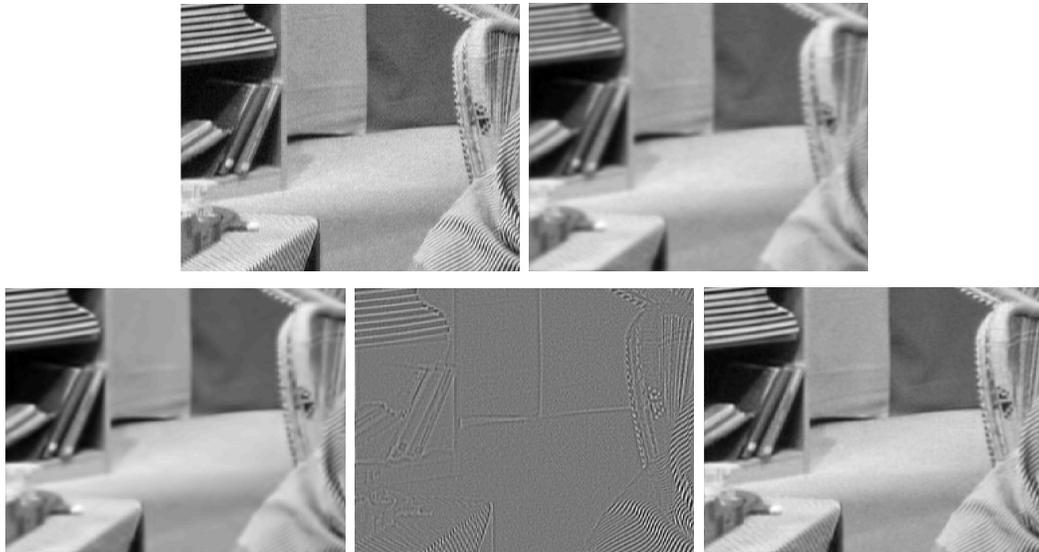


Figure 6: Top from left to right: initial image f , blurred image Kf ; bottom from left to right: deblurred \tilde{u} , deblurred $\tilde{v} + 150$, deblurred $\tilde{u} + \tilde{v}$, computed with Db3 using the iterative approach, $\alpha = 0.2$, $\gamma = 0.001$.

ULB for their hospitality during his stays in Princeton and Brussels.

References

- [1] A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.
- [2] A. Cohen, R. DeVore, P. Petrushev, and H. Xu. Nonlinear Approximation and the Space $BV(\mathbb{R}^2)$. *American Journal of Mathematics*, (121):587–628, 1999.
- [3] R.R. Coifman and D. Donoho. Translation-invariant de-noising. in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, New York, pages 125–150, 1995.
- [4] W. Dahmen. Stability of multiscale transformations. *The Journal of Fourier Analysis and Applications*, 2:341–361, 1996.
- [5] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.
- [6] I. Daubechies. Wavelet transforms and orthonormal wavelet bases. *Proceedings of Symposia in Applied Mathematics*, (47), 1993.
- [7] I. Daubechies, M. Defrise, and C. DeMol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Preprint*, 2003.
- [8] I. Daubechies and G. Teschke. Wavelet based image decomposition by variational functional. *to appear in SPIE proceedings*, 2004.
- [9] M. Defrise and C. DeMol. Inverse imaging with mixed penalties. *Conference proceedings*, 2003.
- [10] M. Defrise and C. DeMol. Linear inverse problems with mixed smoothness and sparsity constraints. *Preprint*, 2003.
- [11] R. DeVore. Nonlinear Approximation. *Acta Numerica*, 7:51–150, 1998.
- [12] R. DeVore, B. Jawerth, and V. Popov. Interpolation of besov spaces. *Trans. Math. Soc.*, 305:397–414, 1988.
- [13] R. DeVore, B. Jawerth, and V. Popov. Compression of wavelet decompositions. *Amer. J. Math.*, 114:737–785, 1992.
- [14] F. C. A. Fernandes, R. v. Spaendonck, M. J. Coates, and S. Burrus. Directional Complex-Wavelet Processing. *Proceedings of Society of Photo-Optical Instrumental Engineers—SPIE2000, Wavelet Applications in Signal Processing VIII*, San Diego., 2000.
- [15] M. Frazier and B. Jawerth. A discrete transform and decompositions of distribution spaces. *J. of Functional Anal.*, 93:34–170, 1990.
- [16] N. Kinsbury. Image processing with complex wavelets. *Phil. Trans. R. Soc. Lond.*, Sept. 1999.

- [17] S. Mallat and S. Zhong. Characterization of Signals from Multiscale Edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7):710–732, July 1992.
- [18] Y. Meyer. Oscillating Patterns in Image Processing and Nonlinear Evolution Equations. *University Lecture Series Volume 22*, AMS, 2002.
- [19] S. Osher, A. Sole, and L. Vese. Image decomposition and restoration using total variation minimization and the H^{-1} norm. Technical Report 02-57, University of California Los Angeles C.A.M., 2002.
- [20] S. Osher and L. Vese. Modeling textures with total variation minimization and oscillating patterns in image processing. Technical Report 02-19, University of California Los Angeles C.A.M., 2002.
- [21] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variations based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [22] I. W. Selesnick. Hilbert transform pairs of wavelet bases. *IEEE Signal Processing Letters*, 8(6):170–173, June 2001.
- [23] H. Triebel. *Interpolation Theory, Function Spaces, Differential Operators*. Verlag der Wissenschaften, Berlin, 1978.