Structure-Texture Image Decomposition – Modeling, Algorithms, and Parameter Selection

Jean-François Aujol, Guy Gilboa, Tony Chan & Stanley Osher

UCLA, Departement of Mathematics, Los Angeles, California 90095,
email: {aujol,gilboa,chan,sjo}@math.ucla.edu

July 28, 2005

Abstract

This paper explores various aspects of the image decomposition problem using modern variational techniques. We aim at splitting an original image f into two components u and v, where u holds the geometrical information and v holds the textural information. The focus of this paper is to study different energy terms and functional spaces that suit various types of textures.

Our modeling uses the total-variation energy for extracting the structural part and one of four of the following norms for the textural part: L^2, G, L^1 and a new tunable norm, suggested here for the first time, based on Gabor functions.

Apart from the broad perspective and our suggestions when each model should be used, the paper contains three specific novelties: first we show that the correlation graph between u and v may serve as an efficient tool to select the splitting parameter, second we propose a new fast algorithm to solve the TV-L^1 minimization problem, and third we introduce the theory and design tools for the TV-Gabor model.

Key-words: Image decomposition, restoration, parameter selection, BV, G, L^1, Hilbert space, projection, total-variation, Gabor functions.

1 Introduction

1.1 Motivation

Decomposing an image into meaningful components is an important and challenging inverse problem in image processing. A first range of models are denoising models: in such models, the image is assumed to have been corrupted by noise, and the processing purpose is to remove the noise. This task can be regarded as a decomposition of the image into signal parts and noise parts. Certain assumptions are taken with respect to the signal and noise, such as the piecewise smooth nature of the image, which enables good approximations of the clean original image.

In modern image-processing, two main successful approaches are usually considered to solve the denoising problem. The first one is based on manipulating the wavelet coefficients of the image [20, 35, 14, 34, 33, 37]. The second one is based on solving nonlinear partial-differential equations (PDE's) associated with the minimization of an energy composed of some norm of the gradient [44, 13, 4, 37, 40, 41].

A related but different problem, which is the main topic of this paper, is the decomposition of an image into its structural and textural parts. The aim of this type of decomposition is harder to formulate explicitly. The general concept is that an image can be regarded as composed of a structural part, corresponding to the main large objects in the image, and a textural part, containing fine scale-details, usually with some periodicity and oscillatory nature. The definition of texture is vague and highly depends on the image scale. A "structure" in one scale, can be regarded as "texture" in another scale. Nevertheless, we will attempt to use various variational models, to decompose an image into meaningful structural and textural parts. Moreover, we will examine the ability to perform the task automatically using the correlation criterion. This criterion is very simple and does not assume any information on the nature or scale of the texture. It works well in simple cases and can aid in finding the right weight between the structural and textural components. In more complicated multi-scale images, more elaborated mechanisms are needed, based on additional information. We will discuss the advantages and drawbacks of the correlation criterion and suggest possible ways for further research to solve this difficult problem.

In this paper, we will focus on image decomposition models based on total variation regularization methods, as originally proposed in [44]. This approach has recently been analyzed in [37], which is the inspiration source of many works [50, 42, 5, 3, 45, 7, 11, 15, 19, 32, 51]. In Section 2 we review the decomposition models that are considered in this paper.

We aim at splitting an original image f into two components u and v, u containing the geometrical information and v the textural information. Our modeling is based on TV regularization approaches: we minimize a functional with two terms, a first one based on the total variation and a second one on a different norm adapted to the texture component. One aim of the paper is to analyze the different structure-texture models and to point out the similarities and differences between the decomposition techniques. In addition, three main contributions are presented:

1. First, we show that the correlation graph between u and v is an efficient tool to select the splitting parameter.

2. Second, we propose a new fast algorithm to solve the TV-L^1 minimization problem.

3. Third, we introduce a new TV-Gabor model which leads us to adaptive frequency and directional image decomposition.

All the algorithms we consider are inspired by the ROF model [44], in the sense that they are all of the generic form *TV-another norm*.

1.2 Outline of the paper

A main purpose of the paper is to know when one should use each model. The organization of the paper is influenced by our final conclusions. The four models can be classified to fit three main types of textures: general oscillating patterns ($TV-L^2$ and $TV-G$), structural textures ($TV-L^1$) and smooth periodic, possibly directional, textures (TV-Gabor).

The paper is organized as follows. In Section 2, the four decomposition models are formulated. In Section 3, we introduce the notations that will be used in the rest of the paper. We briefly review Chambolle's projection algorithm, which is a recent and efficient method to solve the ROF problem [12]. We recall how Chambolle's algorithm can be used to solve the A^2BC model [5]. We also recall the framework of the TV-Hilbert regularization of [8]. In Section 4, we propose a method to compute the decomposition of an image using a correlation criterion, inspired by the work of [39]. In Section 5, we examine general type decompositions using the $TV-L^2$ and $TV-G$ models and relate their parameters in the A^2BC framework (which well approximates $TV-G$). In Section 6, we introduce a new fast and efficient algorithm to solve the $TV-L^1$ minimization problem (5). We carry out the complete mathematical analysis of this new algorithm. The advantages and drawbacks of using the correlation method for parameter tuning to this kind of regularization are presented. In Section 7, we design a family of Hilbert spaces based on Gabor functions. This provides us with a new TV-Gabor model in which one can take advantage of knowledge of both the frequency and the direction of the texture. It is also shown how the correlation criterion can be used to select the regularization parameter. We then conclude the paper in Section 8 with some final remarks and future prospects. In Appendix A.1, we detail the proofs of the mathematical results of Section 6.

2 Four Decomposition Models

From now on, we denote by f the original image to decompose. It is reasonnable and classical to assume that f is defined on a bounded and connected Lipschitz open set Ω (typically Ω is a rectangle), and that f is bounded. Therefore f belongs to $L^\infty(\Omega)$. Since Ω is bounded, f also belongs to $L^2(\Omega)$.

2.1 TV-L^2 (ROF)

Rudin, Osher and Fatemi proposed in [44] a popular denoising algorithm which preserves well the edges of the original image, while removing most of the noise. This algorithm decomposes an image f into a component u belonging to BV and a component v in L^2. In this approach the following functional is being minimized:

$$\inf_{(u,v)\in BV\times L^2/f=u+v} \left(\int |Du| + \lambda \|v\|_{L^2}^2 \right) \qquad (1)$$

where $\int |Du|$ is the total variation of u. For a detailed mathematical study of (1) we refer the reader to [13].

2.2 TV-G (Meyer)

In [37], Meyer suggests a new decomposition model. He proposes the following functional:

$$\inf_{(u,v)\in BV\times G/f=u+v} \left(\int |Du| + \lambda \|v\|_G \right) \qquad (2)$$

where the Banach space G contains signals with large oscillations, and thus in particular textures and noise. We give here the definition of G.

Definition 1 *G is the Banach space composed of distributions f which can be written as*

$$f = \partial_1 g_1 + \partial_2 g_2 = \operatorname{div}(g) \tag{3}$$

with g_1 and g_2 in L^∞. The space G is endowed with the following norm:

$$\begin{aligned} \|v\|_G \quad = \quad & \inf \{ \|g\|_{L^\infty} \ / v = \operatorname{div}(g),\ g = (g_1, g_2), \\ & g_1 \in L^\infty, \quad g_2 \in L^\infty, \\ & |g(x)| = \sqrt{(|g_1|^2 + |g_2|^2)(x)} \Big\} \end{aligned} \tag{4}$$

A function belonging to G may have large oscillations and nevertheless have a small norm. Thus the norm on G is well-adapted to capture the oscillations of a function in an energy minimization method. We refer the reader to [7] for some numerical computations of typical image G norm. In [37], the author did not propose any numerical scheme to compute the decomposition. Vese and Osher [50] were the first to propose a numerical scheme to solve this model using Euler-Lagrange equations based on L^p norms. Aujol et al [6, 5] suggested a different method based on projection (A^2BC) which will be explained in Section 3.3. Notice that an approach based on second order cone programming has recently been proposed in [51].

2.3 TV-L^1

In [2] and [40] it was suggested to replace the L^2 norm in the ROF model by a L^1 norm. The functional to minimize in this case is

$$\inf_{(u,v) \in BV \times L^1 / f = u+v} \left(\int |Du| + \lambda \|v\|_{L^1} \right) \tag{5}$$

Nikolova has showed that the L^1 norm is particularly well suited to remove salt and pepper noise [40]. Comparing to the ROF model (1), this functional does not erode structures, and presents other interesting properties.

This model has recently been studied mathematically in the continuous case in [15]. The authors present interesting quantitative properties of the model related to scale-space and show that geometrical features are better preserved. Numerically, one of the main drawbacks of the model is that, until now, there was no fast algorithm to solve (5). An important contribution of the paper is to address this problem and to propose a fast and efficient algorithm to solve (5). We will study problem (5) in Section 6. A different method based on second order cone programming has recently been proposed in [51].

2.4 TV-Hilbert

Motivated by [44] and [42], the authors of [8] have proposed a generalization of the ROF and OSV models:

$$\inf_{(u \times v) \in BV \times \mathcal{H} / f = u+v} \left\{ \int |Du| + \lambda \|v\|^2_{\mathcal{H}} \right\} \tag{6}$$

where $\mathcal{H}$ is some Hilbert space. In the case when $\mathcal{H} = L^2$, then (6) is the ROF model [44], and when $\mathcal{H} = H^{-1}$ then (6) is the OSV model [42]. By choosing suitably the Hilbert space $\mathcal{H}$, it is possible to compute a frequency and directional adaptive image decomposition, as we will see in Section 7. One of the main contributions of the paper is the designing of a family of Hilbert spaces based on Gabor wavelets for such a purpose.

3 Settings and Previous Projection Algorithms

In this paper all our models are solved numerically by projections algorithms, and not by using the more classical techniques based on Euler-Lagrange equations. Notice that a method based on convex analysis to solve TV models was recently proposed in [18], and another one based on Support Vector Regression in [46] We present Chambolle's projection algorithm, which is a recent method to solve the ROF problem [12]. An important advantage of this algorithm is that there is no need to regularize the TV energy. When using Euler-Lagrange equations to minimize a TV term, one first needs to regularize the functional and consider instead $\int \sqrt{|\nabla u|^2 + \epsilon^2}$. The small parameter ϵ is necessary to prevent numerical instabilities. The main advantage of Chambolle's projection method is that it does not use this additional artificial parameter, and is therefore more faithful to the continuous formulation of the energy. Moreover, in this projection framework, we can easily and rigourously show the convergence of the algorithms towards the minimizers of the functional.

We also recall how Chambolle's algorithm can be used to solve the A^2BC model [5]. We then recall how Chambolle's algorithm has been extended to a larger class of TV-Hilbert functionals in [8]. We begin by introducing the notations that we will use in the rest of the paper.

3.1 Discretization

From now on and through the rest of the paper, we consider the discrete case The image is a two dimension vector of size $N \times N$. We denote by X the Euclidean space $\mathbb{R}^{N \times N}$, and $Y = X \times X$. The space X

will be endowed with the L^2 inner product $(u,v)_{L^2} = \sum_{1\leq i,j\leq N} u_{i,j}v_{i,j}$ and the norm $\|u\|_{L^2} = \sqrt{(u,u)_{L^2}}$. We also set $\|u\|_{L^1} = \sum_{1\leq i,j\leq N} |u_{i,j}|$. To define a discrete total variation, we introduce a discrete version of the gradient operator. If $u \in X$, the gradient ∇u is a vector in Y given by: $(\nabla u)_{i,j} = ((\nabla u)^1_{i,j}, (\nabla u)^2_{i,j})$, with

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N \\ 0 & \text{if } i = N \end{cases}$$

and

$$(\nabla u)^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < N \\ 0 & \text{if } j = N \end{cases} .$$

The discrete total variation of u is then defined by:

$$J(u) = \sum_{1\leq i,j\leq N} |(\nabla u)_{i,j}| \tag{7}$$

We also introduce a discrete version of the divergence operator. We define it by analogy with the continuous setting by $\text{div} = -\nabla^*$ where ∇^* is the adjoint of ∇: that is, for every $p \in Y$ and $u \in X$, $(-\text{div}\, p, u)_{L^2} = (p, \nabla u)_Y$. It is easy to check that:

$$\begin{aligned} (\text{div}\,(p))_{i,j} &= \begin{cases} p^1_{i,j} - p^1_{i-1,j} & \text{if } 1 < i < N \\ p^1_{i,j} & \text{if i=1} \\ -p^1_{i-1,j} & \text{if i=N} \end{cases} \\ &+ \begin{cases} p^2_{i,j} - p^2_{i,j-1} & \text{if } 1 < j < N \\ p^2_{i,j} & \text{if j=1} \\ -p^2_{i,j-1} & \text{if j=N} \end{cases} \end{aligned} \tag{8}$$

From now on, we will use these discrete operators. We are now in position to introduce the discrete version of Meyer's space G.

Definition 2

$$G = \{v \in X \ / \ \exists g \in Y \text{ such that } v = \text{div}\,(g)\} \tag{9}$$

and if $v \in G$:

$$\begin{aligned} \|v\|_G &= \inf \{\|g\|_\infty \ / \ v = \text{div}\,(g), \\ &\qquad g = (g^1, g^2) \in Y, |g_{i,j}| = \sqrt{(g^1_{i,j})^2 + (g^2_{i,j})^2}\Big\} \end{aligned} \tag{10}$$

where $\|g\|_\infty = \max_{i,j} |g_{i,j}|$.

Moreover, we will denote:

$$G_\mu = \{v \in G \ / \ \|v\|_G \leq \mu\} \tag{11}$$

We recall that the Legendre-Fenchel transform of F is given by $F^*(v) = \sup_u (u,v)_{L^2} - F(u)$ (see [22]). The following result is proved in [5]. We see that $J(.)$ (resp.$\|.\|_G$) is the polar of $\|.\|_G$ (resp. $J(.)$).

Proposition 1 *The space G identifies with the following subspace:*

$$X_0 = \{v \in X \ / \ \sum_{i,j} v_{i,j} = 0\} \tag{12}$$

Notice that these results are in discrete. See [3] for the definition of G in the continuous case. We also refer the interested reader to [30] about the relation between the discrete and the continuous Fenchel dual.

3.2 Chambolle's projection algorithm

Since J defined by (7) is homogeneous of degree one (i.e. $J(\lambda u) = \lambda J(u)\ \forall u$ and $\lambda > 0$), it is then standard (see [22]) that J^* is the indicator function of some closed convex set, which turns out to be the set G_1 defined by (11):

$$J^*(v) = \chi_{G_1}(v) = \begin{cases} 0 & \text{if } v \in G_1 \\ +\infty & \text{otherwise} \end{cases} \tag{13}$$

This can be checked out easily (see [12] for details). In [12], the author proposes a nonlinear projection algorithm to minimize the ROF model. The problem is:

$$\inf_{u\in X} \left(J(u) + \frac{1}{2\lambda}\|f-u\|^2_{L^2} \right) \tag{14}$$

We have the following result, which comes from standard convex duality theory [22]:

Proposition 2 *([12]): The solution of (14) is given by: $u = f - P_{G_\lambda}(f)$ where P is the orthogonal projector on G_λ (defined by (11)).*

We use the following algorithm to compute $P_{G_\lambda}(f)$. It indeed amounts to finding:

$$\min \left\{ \|\lambda \text{div}\,(p) - f\|^2_{L^2} : p \ / \ |p_{i,j}| \leq 1 \ \forall i,j = 1,\ldots,N \right\} \tag{15}$$

This problem can be solved by a fixed point method: $p^0 = 0$ and

$$p^{n+1}_{i,j} = \frac{p^n_{i,j} + \tau(\nabla(\text{div}\,(p^n) - f/\lambda))_{i,j}}{1 + \tau|(\nabla(\text{div}\,(p^n) - f/\lambda))_{i,j}|} \tag{16}$$

In [12] is given a sufficient condition ensuring the convergence of the algorithm: it is shown that as long as $\tau \leq 1/8$, then $\lambda \text{div}\,(p^n)$ converges to $P_{G_\lambda}(f)$ as $n \to +\infty$.

3.3 Aujol-Aubert-Blanc-Féraud-Chambolle model (A^2BC)

Inspired from the work by A. Chambolle [12] and by the numerical results of [50], the authors of [5, 6] propose a relevant approach to solve Meyer problem. They

consider the following problem

$$\inf_{(u,v)\in X\times G_\mu} \left(J(u) + \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 \right) \quad (17)$$

where $G_\mu = \{v \in G / \|v\|_G \leq \mu\}$, and $\|v\|_G$ is defined by (10), and $J(u)$ by (7)

The authors of [5] present their model in a discrete framework. See also [3] for a study of this model in a continuous setting, and [10] for an extension to color images. In this paper, we will focus on the A^2BC model to solve Meyer's problem automatically in Section 5. In [5, 6], the authors use Chambolle's projection algorithm [12] to solve (17). We describe their method below.

Minimization: Since J^* is the indicator function of G_1 (see (13)), we can rewrite (17) as

$$\inf_{(u,v)\in X\times X} \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 + J(u) + J^*\left(\frac{v}{\mu}\right) \quad (18)$$

With this formulation, we see the symmetric roles played by u and v. To solve (18), we consider the two following problems:

- v being fixed, we search for u as a solution of:

$$\inf_{u\in X} \left(J(u) + \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 \right) \quad (19)$$

- u being fixed, we search for v as a solution of:

$$\inf_{v\in G_\mu} \|f-u-v\|_{L^2}^2 \quad (20)$$

From Proposition 2, we know that the solution of (19) is given by: $\hat{u} = f - v - P_{G_\alpha}(f-v)$. And the solution of (20) is simply given by: $\hat{v} = P_{G_\mu}(f-u)$.

Algorithm:

1. Initialization:
$$u_0 = v_0 = 0 \quad (21)$$
2. Iterations:
$$v_{n+1} = P_{G_\mu}(f-u_n) \quad (22)$$
$$u_{n+1} = f - v_{n+1} - P_{G_\alpha}(f-v_{n+1}) \quad (23)$$
3. Stopping test: we stop if
$$\max(|u_{n+1}-u_n|, |v_{n+1}-v_n|) \leq \epsilon \quad (24)$$

It is shown in [5] that the sequence (u_n, v_n) given by (21)-(22)-(23) converges to the unique minimizer of problem (17).

Parameters: Algorithm (21)-(22)-(23) needs thus the two parameters α and μ. The parameter α controls the L^2-norm of the residual $f-u-v$. The smaller α is, the smaller the L^2 norm of the residual $f-u-v$ is. The larger μ is, the more v contains information, and therefore the more u is averaged. In fact, the choice of α is easy. One just needs to set it very small. For instance, in all the examples presented hereafter, we have chosen $\alpha = 1$, and found out a maximum norm for $f-u-v$ of about 0.5 (for values ranging from 0 to 255). But the μ parameter is much harder to tune. It controls the G norm of the oscillating component v. In the case of image denoising, a first method to tune μ with respect to the standard deviation of the noise has been proposed in [7]. We will present a way to select μ in the case of image decomposition in Section 5.

3.4 $\mathcal{H}$ Hilbert space

In [8], the authors have considered other spaces to model oscillating patterns. They propose to use a general family of Hilbert spaces that we will consider in Section 7. These Hilbert spaces are defined thanks to an operator K.

K a linear symmetric positive-definite operator from $\mathcal{A}$ to L^2, where $\mathcal{A}$ is either X_0 or L^2 (we recall that X_0 is defined by (12)). In the case when $\mathcal{A} = X_0$, then we extend K to the whole L^2 by setting $K(x) = +\infty$ if $x \in L^2 \backslash X_0$. Notice that with these assumptions, then we can define K^{-1} on $ImK = \{z \in L^2$ such that $\exists x \in \mathcal{A}$ with $z = K(x)\}$.

If f and g are in X_0, then let us define:

$$\langle f, g\rangle_{\mathcal{H}} = \langle f, Kg\rangle_{L^2} \quad (25)$$

This defines a inner product on $X_0 = \{x \in X \ / \ \sum_{i,j} x_{i,j} = 0\}$.

We note that since we only deal here with the discrete case, all the spaces we consider are of finite dimension and are therefore Euclidean spaces.

Examples:

1. When $K = Id$, then $\mathcal{H} = L^2$.
2. When $K = -\Delta$, then $\mathcal{H} = H = \{f \in L^2, \nabla f \in L^2\}$.
3. When $K = -\Delta^{-1}$, then $\mathcal{H} = H^{-1} = (H_0^1)^*$ (see [1] for the definition of H^{-1}).

Remark: In Section 7, we will assume $\mathcal{A} = L^2$, i.e. that K is positive-definite on L^2.

3.5 TV-Hilbert regularization model

The model studied in [8] is the following:

$$\inf_u \left(J(u) + \frac{\lambda}{2}\|f-u\|_{\mathcal{H}}^2 \right) \tag{26}$$

In [8], the authors give some mathematical results about this problem. In particular, they show the existence and uniqueness of a solution for (26). They also propose a modification of Chambolles's projection algorithm [12] to compute the solution of problem (26):

$$p^0 = 0 \tag{27}$$

and

$$p_{i,j}^{n+1} = \frac{p_{i,j}^n + \tau(\nabla(K^{-1}\mathrm{div}\,(p^n) - \lambda f))_{i,j}}{1+\tau|(\nabla(K^{-1}\mathrm{div}\,(p^n) - \lambda f))_{i,j}|} \tag{28}$$

Theorem 1 *If* $\tau \leq \frac{1}{8\|K^{-1}\|_{L^2}}$, *then* $\frac{1}{\lambda}K^{-1}\mathrm{div}\,p^n \to \hat{v}$ *as* $n \to \infty$, *and* $f - \frac{1}{\lambda}K^{-1}\mathrm{div}\,p^n \to \hat{u}$ *as* $n \to \infty$, *where* $\hat{u}$ *is the solution of problem (26) and* $\hat{v} = f - \hat{u}$.

In [8], the authors apply their framework to solve the OSV model [42] (i.e. when $\mathcal{H} = H^{-1}$), and they study the problem of image denoising. In this paper, we intend to use (26) to carry out frequency and directional adaptive image decomposition. Indeed, by choosing the kernel K in a suitable way, we can emphasize the weight of some frequencies and directions. We will address this problem in Section 7.

Now that we have introduced the notations and presented some of the previous works, we present a general criterion based on correlation to select the regularization parameter in the different models that we will consider.

4 The Correlation Tool for Selecting the Balance between the Energies

In this section, we propose a method to select the weight parameter for a proper decomposition of an image. The authors are not aware of any suggested method on how to choose the value of λ for decomposition. Therefore we first discuss shortly the solutions at present that are used for denoising and explain the difficulties that arise in decomposition.

For the denoising problem, one often assumes that the variance of the noise σ^2 is known *a-priori* or can be well estimated from the image. As the v part in the denoising case should contain mostly noise, a natural condition is to select λ such that the variance of v is equal to that of the noise, that is $\mathrm{var}(v) = \sigma^2$. Such a method was used in [44] in the constrained ROF model, and this principle dates back to Morozov [38] in regularization theory. A modern approach, suggested recently in [27], is to try to optimize a criterion, such as the Signal-to-Noise Ratio (SNR). It was shown that this method can achieve better results than the constrained formulation, in terms of SNR and visually, for a wide class of images. This method also relies on an estimation of the noise variance.

Both of the above approaches cannot be applied for finding λ in decomposition. Here we do not know of a good way to estimate the texture variance, also there is no performance criterion like the SNR, which can be optimized. Therefore we should resort to a different approach.

Our approach follows the work of Mrazek-Navara [39], used for finding the stopping time for denoising with nonlinear diffusions. The method relies on a correlation criterion and assumes no knowledge of noise variance. As shown in [27], its performance is inferior to the SNR-based method of [27] and to an analogue of the variance condition for diffusions. For decomposition, however, the approach of [39], adopted for the variational framework, may be a good basic way for the selection of λ.

In this paper the general decomposition framework is of the form:

$$E_{Structure}(u) + \lambda E_{Texture}(v), \quad f = u + v, \tag{29}$$

where u and v minimize the above total energy. Our goal is to find the right balance between the energy terms, or the value of λ, which produces a meaningful structure-texture decomposition.

Let us define first the (empirical) notions of mean, variance and covariance in the discrete setting of $N \times N$ pixels image. The mean is

$$\bar{q} \doteq \frac{1}{N^2} \sum_{1 \leq i,j \leq N} q_{i,j},$$

the variance is

$$V(q) \doteq \frac{1}{N^2} \sum_{1 \leq i,j \leq N} (q_{i,j} - \bar{q})^2,$$

and the covariance is

$$\mathrm{cov}(q,r) \doteq \frac{1}{N^2} \sum_{1 \leq i,j \leq N} (q_{i,j} - \bar{q})(r_{i,j} - \bar{r}).$$

We would like to have a measure that defines orthogonality between two signals and is not biased by the

magnitude (or variance) of the signals. A standard measure in statistics is the correlation, which is the covariance normalized by the standard deviations of each signal:

$$\mathrm{corr}(q,r) \doteq \frac{\mathrm{cov}(q,r)}{\sqrt{V(q)V(r)}}.$$

By the Cauchy-Schwarz inequality it is not hard to see that $\mathrm{cov}(q,r) \leq \sqrt{V(q)V(r)}$ and therefore $|\mathrm{corr}(q,r)| \leq 1$. The upper bound 1 (completely *correlated*) is reached for signals which are the same, up to an additive constant and up to a positive multiplicative constant. The lower bound -1 (completely *anti-correlated*) is reached for similar signals but with a negative multiplicative constant relation. When the correlation is 0 we refer to the two signals as *not correlated*. This is a necessary condition (but not a sufficient one) for statistical independence. It often implies that the signals can be viewed as produced by different "generators" or models.

To guide the parameter selection of a decomposition we use the following assumption:

Assumption: *The texture and the structure components of an image are not correlated.*

This assumption can be relaxed by stating that the correlation of the components is very low. Let us define the pair (u_λ, v_λ) as the one minimizing (29) for a specific λ. As proved in [37] for the $TV - L^2$ model (and in [24] for any convex structure energy term with L^2), we have $\mathrm{cov}(u_\lambda, v_\lambda) \geq 0$ for any non-negative λ and therefore

$$0 \leq \mathrm{corr}(u_\lambda, v_\lambda) \leq 1, \ \forall \lambda \geq 0. \tag{30}$$

This means that one should not worry about negative correlation values. Note that positive correlation is guaranteed in the $TV - L^2$ case. As we will later see, in the $TV - L^1$ case we may have negative correlations, and should therefore be more careful.

Following the above assumption and the fact that the correlation is non-negative, to find the right parameter λ, we are led to consider the following problem:

$$\lambda^* = \mathrm{argmin}_\lambda \left(\mathrm{corr}(u_\lambda, v_\lambda)\right). \tag{31}$$

In practice, one generates a scale-space using the parameter λ (in our formulation, smaller λ means more smoothing of u) and selects the parameter λ^* as the first local minimum of the correlation function between the structural part u and the oscillating part v. See also [24, 25, 28, 26, 39, 8] for related approaches.

This selection method can be very effective in simple cases with very clear distinction between texture and structure. In these cases $\mathrm{corr}(u,v)$ behaves smoothly, reaches a minimum approximately at the point where the texture is completely smoothed out from u, and then increases, as more of the structure gets into the v part. See Figures 1 to 5 in the next section for some numerical examples. The graphs of $\mathrm{corr}(u,v)$ in the $TV-L^2$ case behave quite as expected, and the selected parameter lead to a good decomposition. We will make more comments about the numerical results in the next section.

For more complicated images, there are textures and structures of different scales and the distinction between them is not obvious. In terms of correlation, there is no more a single minimum and the function may oscillate.

As a first approximation of a decomposition with a single scalar parameter, we suggest to choose λ after the first local minimum of the correlation is reached. In some cases, a sharp change in the correlation is also a good indicator: after the correlation sharply drops or before a sharp rise. At this stage we cannot claim a fully automatic mechanism for the parameter selection that always works, but rather a highly relevant measurement that should be taken into consideration in future development of automatic decompositions.

5 $TV - L^2$ and $TV - G$ Regularizations

In this section, we first show how we can use the correlation tool to select the parameter in the $TV - L^2$ regularization model. We then show how we can extend this method to the $TV - G$ model.

5.1 Parameter selection for the $TV - L^2$ model

Let us first recall here the $TV - L^2$ problem [44]:

$$\inf_{u \in X} \left(J(u) + \frac{1}{2\beta} \|f - u\|_{L^2}^2 \right) \tag{32}$$

We denote by (u_β, v_β) the solution of (32). This regularization model has encountered a large success in image denoising. One of the main reason of this success is that the total variation regularization preserve the edges of the restored image. It is straightforward to apply the correlation criterion of Section 4 to select the parameter in the $TV - L^2$ model.

5.2 Parameter selection for the $TV - G$ model

We focus here on the A^2BC model [5], which is a very good approximation. We show how we can use the correlation criterion for the ROF model [44] to carry out automatic image decomposition with the A^2BC model. A first approach would be to consider the correlation between u and v computed with the A^2BC algorithm. We have rejected this approach because of computation time: indeed, to compute an accurate solution with the A^2BC algorithm is about ten times slower than the classical $TV - L^2$ minimization approach. We have decided instead to use the mathematical connections between the ROF model and the A^2BC algorithm to select the parameter in a much faster way.

To this end, we first need to give some mathematical properties of the A^2BC model, (17), which is a way to solve Meyer's problem. As we have said in Section 3, the parameter α in (17) is set to a fixed small value ($\alpha = 1$ in our numerical examples). The difficulty is to tune the μ parameter. We intend here to propose a method to compute automatically the parameter μ. The idea is to use the method proposed for the ROF model in Section 5.1 (which is a straightforward application of the general method presented in Section 4). By choosing β as the first minimum of the function $\beta \mapsto \text{corr}(u_\beta, v_\beta)$ (where u_β is the solution of the ROF problem (32) and $v_\beta = f - u_\beta$), we have an automatic algorithm to compute the right parameter β for (32). All we need to do then is to relate the parameter β in (32) to μ in the A^2BC model (17).

5.2.1 Relating β to μ

In [37], Meyer introduced the G norm to analyze the mathematical properties of the ROF model. As noticed in [48], one of the main results of [37] happens to be a straightforward corollary of Proposition 2:

Corollary 1 *Let us denote by u_β the solution of (32), by $v_\beta = f - u_\beta$, and by $\bar{f}$ the mean of f.*

- *If $\|f - \bar{f}\|_G \geq \beta$, then $\|v_\beta\|_G = \|f - u_\beta\|_G = \beta$.*
- *If $\|f - \bar{f}\|_G \leq \beta$, then $u_\beta = \bar{f}$.*

As we can see, the behavior of the ROF model is closely related to the G norm of the initial data f.

Lemma 1 *The parameter β computed in Section 4 is such that $\|v_\beta\|_G = \beta$.*

Proof. Let us denote $\beta_{\max} = \|f - \bar{f}\|_G$. It is easy to show that if $\beta \in (0, \beta_{\max})$, then $\text{corr}(u_\beta, v_\beta)$ remains bounded. From Corollary 1, we get that if $\beta \geq \beta_{\max}$, then $u_\beta = \bar{f}$ and $v_\beta = f - \bar{f}$. Therefore the first local minimum of the correlation is such that $\beta \leq \beta_{\max}$. We then conclude thanks to Corollary 1. □

Thanks to Section 5.1, we know how to compute automatically the decomposition of an original image with the ROF model. And thanks to Lemma 1, we also know the G norm of the v component we get with the ROF model, i.e. $\|v\|_G = \beta$. As we have explained in the introduction, Meyer's idea is to replace the L^2 norm in the ROF model (32) by the G norm. The G norm is better suited to capture oscillating patterns, such as textures, than the L^2 norm (as it is numerically shown in [7]). Therefore, a possible improvement of the algorithm of Section 5.1 is to compute Meyer's decomposition under the constraint that $\|v\|_G = \beta$. Since the G norm is a better choice to capture the texture part of an image [37, 3, 7], this would indeed gives a better decomposition result than the ROF model.

This naturally leads us to consider the A^2BC model (17) with

$$\mu = \beta \tag{33}$$

Indeed, with such a parameter, the v component computed with the A^2BC model is such that $\|v\|_G \leq \beta$. And we prove in the following subsection that in fact we have $\|v\|_G = \beta$.

5.2.2 Some mathematical results about the A^2BC model

The functional to minimize in (17) is the following:

$$F(u, v) = J(u) + J^*\left(\frac{v}{\mu}\right) + \frac{1}{2\alpha}\|f - u - v\|_{L^2}^2 \tag{34}$$

The following Lemma is proved in [5]:

Lemma 2 *There exists a unique couple $(\hat{u}, \hat{v}) \in X \times G_\mu$ minimizing F on $X \times X$.*

From now on, let us denote by $(\hat{u}, \hat{v})$ the unique solution of the A^2BC problem (17). The next result will help to see the connection between the parameter β in the ROF model and the parameter μ in the A^2BC algorithm:

Proposition 3 *The following alternative holds:*

- *If $\|f - \bar{f}\|_G \leq \mu$, then $\hat{v} = f - \bar{f}$.*
- *If $\|f - \bar{f}\|_G \geq \mu$, then $\|\hat{v}\|_G = \mu$.*

Proof. Let us first remark that $F(u, v) \geq 0$ for all (u, v) in $X \times X$. Moreover, if we assume that $\|f - \bar{f}\|_G \leq \mu$, we have $F(\bar{f}, f - \bar{f}) = 0$, which means that $(\bar{f}, f - \bar{f})$ is a minimizer of F. We then get the first point of Proposition 3 thanks to the uniqueness result of Lemma 2.

We now turn our attention to the second point of Proposition 3. We therefore assume that $\|f-\bar{f}\|_G \geq \mu$. Let us consider the following function defined on $X\times X$:

$$H(u,v) = J(u) + \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 \qquad (35)$$

H is a proper convex continuous function defined on $X\times X$. There exists therefore $(\tilde{u},\tilde{v})$ in $X\times G_\mu$ such that $(\tilde{u},\tilde{v})$ is a minimizer of H on $X\times G_\mu$. Let us remark that $H(\bar{f}, f-\bar{f}) = 0$. We then consider the function $g:\ t\mapsto \|t\tilde{v}+(1-t)(f-\bar{f})\|_G$. g is a continuous function on $[0,1]$. Moreover, we have $g(0) = \|f-\bar{f}\|_G \geq \mu$ and $g(1) = \|\tilde{v}\|_G \leq \mu$. There exists thus $\check{t}$ in $[0,1]$ such that $g(\check{t}) = \|\check{t}\tilde{v} + (1-\check{t})(f-\bar{f})\|_G = \mu$. Let us denote by $\check{v} = \check{t}\tilde{v} + (1-\check{t})(f-\bar{f})$ and $\check{u} = \check{t}\tilde{u} + (1-\check{t})\bar{f}$. Since H is a convex function, we get that $H(\check{u},\check{v}) \leq \check{t}H(\tilde{u},\tilde{v}) + (1-\check{t})H(\bar{f}, f-\bar{f}) \leq H(\tilde{u},\tilde{v})$. We therefore deduce that $(\check{u},\check{v})$ is a minimizer of H on $X\times G_\mu$. Since H and F coincide on $X\times G_\mu$, we get that $(\check{u},\check{v})$ is a minimizer of F on $X\times X$. From Lemma 2, we then conclude that $(\check{u},\check{v}) = (\hat{u},\hat{v})$ the unique minimizer of F on $X\times X$, and $\|\hat{v}\|_G = \|\check{v}\|_G = \mu$. $\square$

From Lemma 1 and Corollary 1, we know that $\|f-\bar{f}\|_G \geq \beta$. And from (33), we have $\beta = \mu$. From Proposition 3, we thus deduce that $\hat{v}$, the v component we get with the A^2BC algorithm, is such that $\|\hat{v}\|_G = \mu$. This new v component has therefore the same G norm as the one of the v component (v_β) computed with the ROF model in Section 5.1. But since the G norm is better at capturing the oscillating patterns then the L^2 norm, this new decomposition is more accurate than the previous one.

This analysis is confirmed by the numerical results we get in the next subsection.

5.3 Numerical results

Let us first summarize the method we propose to compute the decomposition into geometry and texture with the A^2BC model.

Automatic algorithm for the A^2BC model:

1. Set $\alpha = 1$ in (17).
2. Compute β as the first minimum of the function $\beta \mapsto \mathrm{corr}(u_\beta, v_\beta)$ (where u_β is the solution of the ROF problem (32) and $v_\beta = f - u_\beta$).
3. Set $\mu = \beta$ in (17).
4. Compute the decomposition with the algorithm (21)-(22)-(23).

We show some numerical results in Figures 1-5 of $TV-L^2$ and $TV-G$ decompositions. As expected, the results obtained with the A^2BC algorithm are slightly better than the ones obtained with the ROF model. For instance, on Figure 1, one can check that the square is less eroded with Meyer's G norm (and in this case, the square is a geometrical feature and should remain in the u component). On Figure 5, one sees that the leg of the table appears much more in the v component with the ROF model than with the A^2BC algorithm. In general, the ROF model already does a good job, and the A^2BC algorithm seems to bring a small improvement. Notice that we do not claim that we compute the best possible results (see [50, 42, 5, 7] for instance where the parameters are tuned manually): what we claim is that our parameter selection method leads to a visually good result (for both models).

Detailed explanation on the correlation graph: In these experiments the correlation $\mathrm{corr}(u,v)$ of 50 values of λ is plotted. We initially set $\lambda^0 = 1$ and reduced each time the value by a factor of 0.9 such that $\lambda^{n+1} = 0.9\lambda^n$. To solve the minimization problem for λ^{n+1} we initialized with the solution obtained for λ^n, and therefore the convergence is quite fast. Also note that in practice one needs not compute the whole graph and can stop when the first local minimum is reached. One may also use courser λ resolutions to save some computational efforts. Note that the correlation graph finds well the right splitting parameter in Figs. 1,2 and even in the more complex Barbara image, Figs. 3-5. In these cases a fully automatic decomposition is possible. In all the correlation graphs the splitting point chosen by our automatic algorithm is marked with "x".

Now that we have introduced a method to automatically compute the μ parameter in (17), that is to automatically compute Meyer's decomposition, we turn our attention to another interesting, more geometric, decomposition model.

6 TV-L^1 Regularization

Let us first recall the model studied in [40]:

$$\inf_u \{J(u) + \lambda\|f-u\|_{L^1}\} \qquad (36)$$

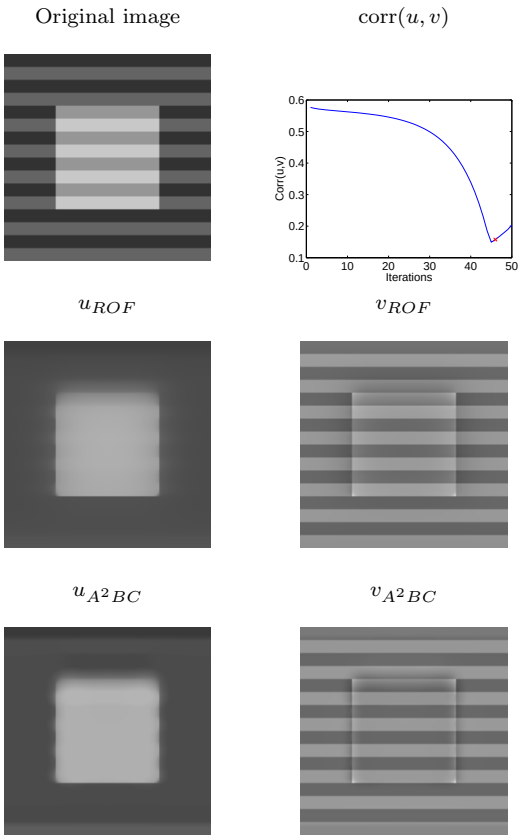


Figure 1: A simple example

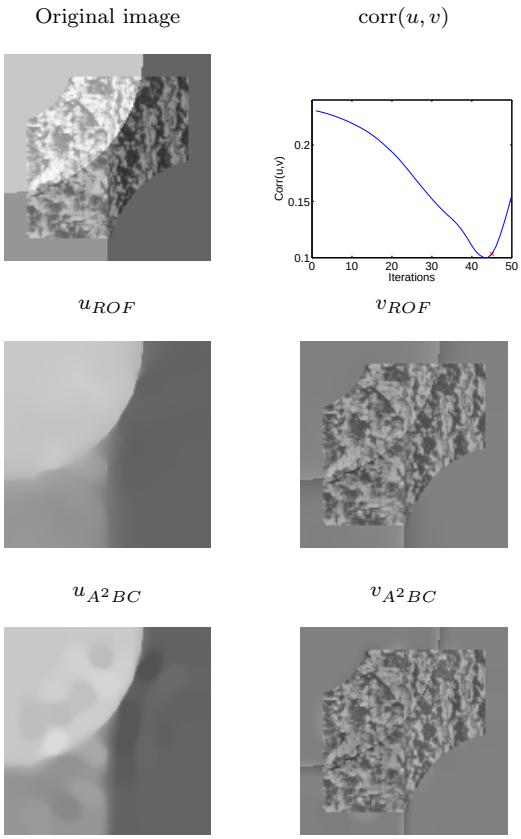


Figure 2: A synthetic image

6.1 A fast algorithm for TV-L^1 regularization

In this section, we introduce a new fast and efficient algorithm to solve the $TV - L^1$ minimization problem (5). We carry out the complete mathematical analysis of this new algorithm. We can then adapt the correlation method for parameter tuning to this kind of regularization.

As we have done previously for the ROF model, we want to derive an automatic algorithm to compute the parameter λ automatically. Our idea is to use the correlation assumption as in Section 4. To this end, we first need to propose a fast algorithm to minimize (36). Indeed, the algorithm used for instance in [15] is a very slow algorithm: the authors first regularize the functional by considering the approximated problem:

$$\inf_u \left\{ \int \sqrt{|\nabla u|^2 + \epsilon_1^2} + \lambda \int \sqrt{(f-u)^2 + \epsilon_2^2} \right\} \quad (37)$$

They compute the solution of this new problem by solving the associated Euler-Lagrange equation.

In [40], the author solves the problem:

$$\inf_u \left\{ \int \sqrt{|\nabla u|^2 + \epsilon^2} + \lambda \|f-u\|_{L^1} \right\} \quad (38)$$

The author proposes a relaxation algorithm to compute the solution, but this is also a slow algorithm. Notice that in this case there may be several possible solutions.

We mention also the very recent work [51] where the authors minimize (38), for $\epsilon = 0$, with an algorithm based on second order cone programming.

6.1.1 A new functional

We remind the reader that in this paper we only consider the discrete case.

We propose here another possible regularization of (36). We consider the functional:

$$\inf_{u,v} \left\{ J(u) + \frac{1}{2\alpha} \|f-u-v\|_{L^2}^2 + \lambda \|v\|_{L^1} \right\} \quad (39)$$

The parameter α is small so that we almost have $f = u + v$.

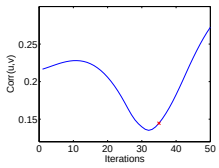


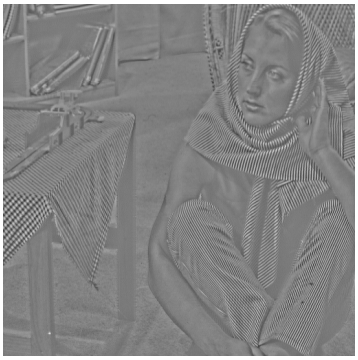
Figure 3: Barbara image and $TV - L^2$ correlation graph.

u_{ROF}

u_{A^2BC}

Figure 4: u component of $TV-L^2$ and $TV-G$ decompositions of the Barbara image (the $TV-G$ decomposition is approximated with the A^2BC algorithm)

v_{ROF}



v_{A^2BC}

Figure 5: v component of $TV-L^2$ and $TV-G$ decompositions of the Barbara image (the $TV-G$ decomposition is approximated with the A^2BC algorithm).

Proposition 4 *β being a positive parameter, the solution $\hat{v}$ of the problem*

$$\inf_v \left\{ \frac{1}{2\beta}\|g-v\|_{L^2}^2 + \|v\|_{L^1} \right\} \tag{40}$$

is given by:

$$v_{i,j} = \begin{cases} g_{i,j} - \beta & \text{if } g_{i,j} \geq \beta \\ 0 & \text{if } |g_{i,j}| \leq \beta \\ g_{i,j} + \beta & \text{if } g_{i,j} \leq -\beta \end{cases} \tag{41}$$

We will write $v = ST(g,\beta)$, i.e. v is the Soft Thresholding of g with level of threshold β.

Proof. The proof is the same as the one proposed in [14] (page 323) in the case of Wavelet Soft Thresholding. It is just a simple 1D minimization problem, since all the equations are independent, and the computation is straightforward. □

Let us now look at the minimization of (39). Since the functional is convex, a natural way to compute the

solution is to minimize with respect to each of the variables separately, and to iterate until convergence as in the A^2BC model for instance. See also [16, 17] for a general approach of such minimization problems. We therefore consider the two following problems:

- v being fixed, we search for u as a solution of:

$$\inf_u \left(J(u) + \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 \right) \tag{42}$$

- u being fixed, we search for v as a solution of:

$$\inf_v \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 + \lambda\|v\|_{L^1} \tag{43}$$

From Proposition 2, we know that the solution of (42) is given by: $\hat{u} = f - v - P_{G_\alpha}(f-v)$. And from Proposition 4, the solution of (43) is given by: $\hat{v} = ST(f-u, \alpha\lambda)$.

It is possible to show as in [5] (for the A^2BC model) that iterating these two minimizations is a way to compute the solution of problem (39). The main advantage of this new algorithm is that instead of the two regularization parameters ϵ_1 and ϵ_2 used in [15], here we only have one regularization parameter λ. Moreover, this new algorithm seems to be faster.

6.1.2 A thresholding algorithm

To increase the speed of the previous algorithm, we propose a slight modification of problem (39). We consider the new functional:

$$\inf_{u,v} \left\{ \|u\|_{\dot{B}^1_{1,1}} + \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 + \lambda\|v\|_{L^1} \right\} \tag{44}$$

where $\dot{B}^1_{1,1}$ is the usual homogeneous Besov space [37, 14, 7].

Although we consider the discrete case, we give here the definition of $\dot{B}^1_{1,1}$ in the continuous case for the sake of clarity.

Definition 3 *Let $\psi_{j,k}$ an orthonormal base composed of smooth and compactly supported wavelets. $\dot{B}^1_{1,1}$ is a subspace of $L^2(\mathbb{R}^2)$, and a function f belongs to $\dot{B}^1_{1,1}$ if and only if: $\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}^2}|c_{j,k}| < +\infty$, where $c_{j,k}$ are the wavelet coefficients of f.*

In this paper, since we want to approximate $J(u)$ by $\|u\|_{\dot{B}^1_{1,1}}$, we only consider the case of the *Haar* wavelet. It is proved in [47] that in 1D, total variation minimization is equivalent to wavelet soft thresholding (in the case of the Haar wavelet with one level of decomposition). However, the two regularization spaces (BV and $\dot{B}^1_{1,1}$) are different. In particular, characteristic functions of sets with finite perimeter belong to BV but are not in $\dot{B}^1_{1,1}$. This is the reason why it can be expected that the edges of the original image f are better put in the geometrical component u with model (39) than with (44).

Let us now look at the minimization of (44). We adopt the same strategy as for solving (39), that is we minimize with respect to each of the variables separately. We therefore consider the two following problems:

- v being fixed, we search for u as a solution of:

$$\inf_u \left(\|u\|_{\dot{B}^1_{1,1}} + \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 \right) \tag{45}$$

- u being fixed, we search for v as a solution of:

$$\inf_v \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 + \lambda\|v\|_{L^1} \tag{46}$$

From [14], we know that the solution of (45) is given by: $\hat{u} = WST(f-v, \alpha)$, where $WST(f-v,\alpha)$ stands for the Wavelet Soft Thresholding of $f-v$ with threshold α [37, 7]. And from Proposition 4, the solution of (46) is given by: $\hat{v} = ST(f-u, \alpha\lambda)$, where $ST(f-u,\alpha\lambda)$ stands for the Soft Thresholding of $f-u$ with threshold $\alpha\lambda$

The advantage for having replaced $J(u)$ by $\|u\|_{\dot{B}^1_{1,1}}$ is that now, to minimize the new functional (44), we just need to iterate thresholding schemes. This is why the following algorithm is a very fast one (much faster than the one used in [15] for instance).

Algorithm:

1. Initialization:

$$u_0 = v_0 = 0 \tag{47}$$

2. Iterations:

$$v_{n+1} = ST(f-u_n, \alpha\lambda) \tag{48}$$

$$u_{n+1} = WST(f-v_{n+1}, \alpha) \tag{49}$$

3. Stopping test: we stop if

$$\max(|u_{n+1}-u_n|, |v_{n+1}-v_n|) \leq \epsilon \tag{50}$$

6.1.3 Mathematical analysis

We now show some mathematical results about our new model, and we prove the convergence of the algorithm. We will use the notation:

$$M(u,v) = \|u\|_{\dot{B}^1_{1,1}} + \frac{1}{2\alpha}\|f-u-v\|^2_{L^2} + \lambda\|v\|_{L^1} \quad (51)$$

Theorem 2 *Problem (44) admits a unique solution* $(\tilde{u},\tilde{v})$ *in* $(X \times X)$.

Proof. See Appendix A.1. □

The next result is a consequence of Theorem 2:

Proposition 5 *The sequence* (u_n, v_n) *built in (47)-(48)-(49) converges to the unique minimizer of problem (44).*

Proof. See Appendix A.2. □

The next result shows that when α goes to 0, then the solution of problem (44) goes to a solution of the problem:

$$\inf_u \left(\|u\|_{\dot{B}^1_{1,1}} + \lambda\|f-u\|_{L^1} \right) \quad (52)$$

Proposition 6 *Let us fix* $\lambda > 0$ *in (52). We consider* α_n *a decreasing sequence in* $\mathbb{R}^*_+$ *such that* $\alpha_n \to 0$. *Let us denote by* $(u_{\alpha_n}, v_{\alpha_n})$ *the solution of problem (44). Then the sequence* $(u_{\alpha_n}, v_{\alpha_n})$ *is bounded, and any cluster point is of the form* $(u_0, f-u_0)$ *with* u_0 *solution of problem (52).*

Proof. See Appendix A.2. □

Remark: It is easy to show that problem (52) has a solution (the functional is convex and coercive). In the case when problem (52) has a unique solution u_0, then the sequence $(u_{\alpha_n}, v_{\alpha_n})$ converges to $(u_0, f-u_0)$.

6.2 Numerical results

A main difference with the classical $TV-L^2$ approach [44] is that with the $TV-L^1$ model, the v component is not constrained to be of zero mean (numerical experiments show that indeed the mean value changes for different values of λ and is not necessarily close to zero).

All the numerical results shown on Figures 6 to 8 have been obtained with the algorithm (47)-(48)-(49), the parameter λ being computed automatically. The parameter α is set to 1 in all our experiments. The maximal absolute values of the computed residuals $f-u-v$ are always smaller than 1 (and the values of the images rank from 0 to 255). This means that the residual energy term, needed mainly for numerical and theoretical reasons (uniqueness), does not affect much the model and the decomposition results.

Remark about parameter selection: To choose the parameter λ, we consider the correlation graph as in Section 4. The difference is that in this case we are not interested in a local minimum of the graph, but in a large variation. This is related to the non-smooth behavior of $TV-L^1$ regularization, as pointed out in [15]. We should remark that the correlation can attain also negative values, unlike the $TV-L^2$ case. If one is interested in decorrelation between u and v, one should seek values close to zero and not minimal ones.

Figure 6 shows an example of removing salt and pepper noise. This relates decomposition to the denoising problem, where in this case the structural part is the clean image and the noise can be regarded as a form of texture. It has been shown in [40] that the L^1 term is particularly well suited to remove such a noise. This is due to the close connections between the L^1 norm and the median operator. In this simple case, the restoration is almost perfect. In Figure 6, second row, the decomposition at the 6th iteration is shown, right after the significant correlation change. Most of the noise is already filtered, but it is better to assume a steadier correlation state, such as at the 10th iteration, depicted in the bottom row. In Figures 7 a decomposition of non-geometric texture is shown. The result is relatively good, though somewhat different than the decomposition of the same image by $TV-L^2$ and $TV-G$ (see Fig. 2). The structural part is less eroded and edges are strong. However, the rounded top left part is not recovered well, and tends to be blocky.

In the case of Figure 8, the decomposition is exact (in this case, the maximum of the absolute of the residual $f-u-v$ is equal to 0.001), and the result is perfect (it is clearly better than the results of Figure 1). The L^1 norm seems to be particularly well suited to capture non smooth textures. Notice however that this image is particularly well suited for the Haar wavelet.

We present, in Figure 9, the decomposition results obtained with the algorithm of [51], which exactly solves (36) as a second-order cone program. We see that both algorithms give similar results for salt and peper noise removal as well as for structure/texture separation.

In this section, we have mainly considered non smooth textures. On the contrary, we will consider smooth textures in the next section.

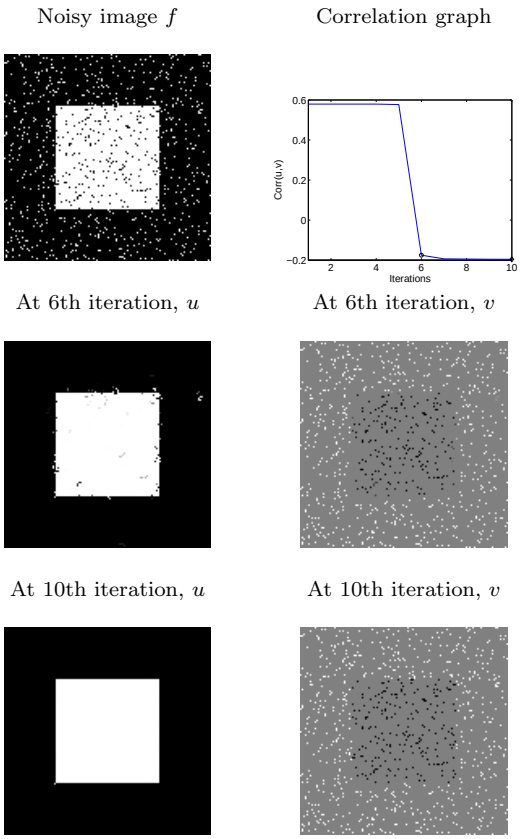


Figure 6: Removing salt and pepper noise (algorithm (47)-(48)-(49)).

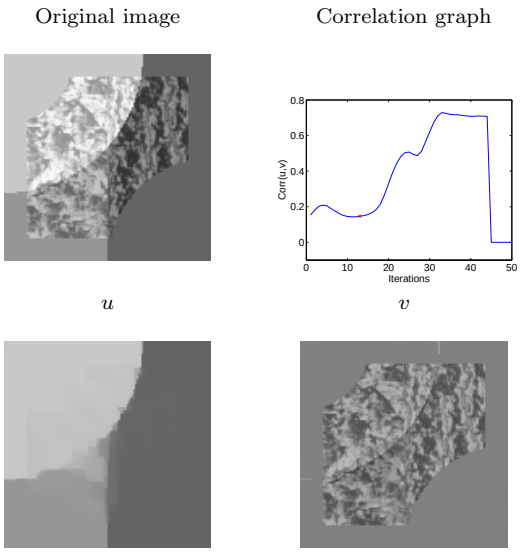


Figure 7: Approximation of the TV-L1 decomposition of non-geometric texture (algorithm (47)-(48)-(49))

.

7 TV-Gabor Regularization

In this section, we design a family of Hilbert spaces based on Gabor wavelets [36]. Gabor functions, proposed by [23], have been found to be very useful in texture processing applications, e.g. [21, 31], and to have close relations with the human-visual system [43]. The Gabor wavelets were also defined by Zibulski and Zeevi in the context of Multiwindow Gabor frames [52]. We introduce a new TV-Gabor model in which one can take advantage of *a-priori* knowledge of both the frequency and the direction of the textures of interest. We show how the correlation criterion can be used also in this case to select the regularization parameter.

7.1 Introduction

Let us first recall the model studied in [8]:

$$\inf_u \left(J(u) + \frac{\lambda}{2}\|f-u\|^2_{\mathcal{H}} \right) \tag{53}$$

In [8], the authors apply their framework to solve the OSV model [42] (i.e. when $\mathcal{H} = H^{-1}$), and they study the problem of image denoising. Here, we intend to use (53) to carry out frequency and directional adaptive image decomposition. Indeed, by choosing the kernel K in a suitable way, we can emphasize the weight of some frequencies and some directions. Notice that, even though K is a linear filter, solving (53) does not amount to linear filtering due to the non linear term $J(u)$. It is well known in image processing that linear filtering cannot preserve edges in an image, but thanks to the total variation term (53) does not suffer from this drawback. To construct the "texture-norm" we use Gabor wavelets.

The projection algorithm proposed in [8] to solve (53) is given by (27)-(28) (in Section 3). In fact, one needs to use K^{-1} and not K to solve (53) with this algorithm. It is therefore easier to construct K^{-1} (so that K has some good properties, but without computing K explicitly). K needs to be a non negative symmetric linear operator. Here we even assume that K is positive-definite. This implies that K^{-1} is also a symmetric positive linear operator.

Remark on a possible alternative construction: K being a positive-definite symmetric operator, there exists a unique positive-definite symmetric linear operator, denoted by $\sqrt{K}$, such that $\sqrt{K}^2 = K$. In particular, we have $\|f-u\|^2_{\mathcal{H}} = \langle f-u, K(f-u)\rangle_{L^2} =$

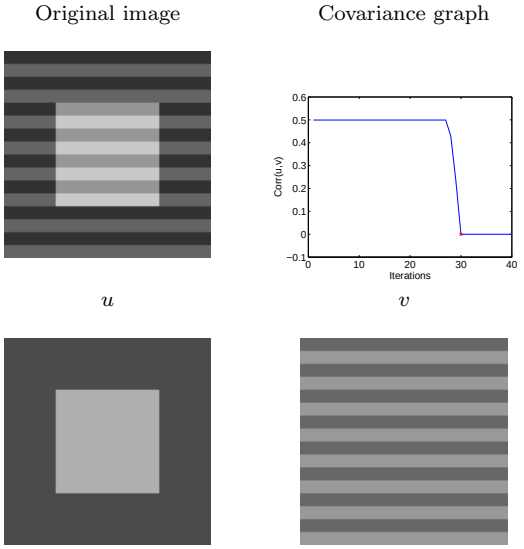


Figure 8: Approximation of the TV-L1 decomposition of non-geometric texture (with non-invariant Haar wavelet soft-thresholding)) with algorithm (47)-(48)-(49))

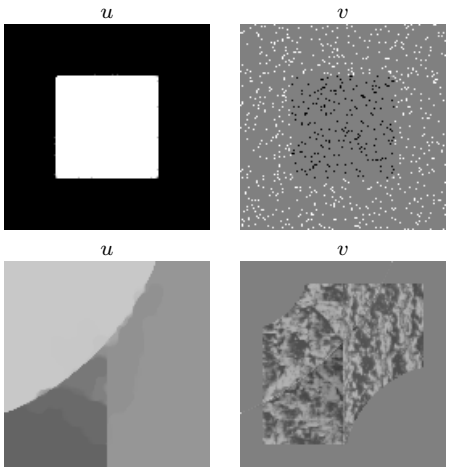


Figure 9: TV-L1 decomposition with the algorithm of [51]. First raw: restoration of the image of Figure 6; second raw: decomposition of the image of Figure 7.

$\|\sqrt{K}(f-u)\|^2_{L^2}$. We can then rewrite problem (53) as:

$$\inf_u \left(J(u) + \frac{\lambda}{2}\|\sqrt{K}(f-u)\|^2_{L^2} \right) \tag{54}$$

In fact, instead of K^{-1}, it also may be interesting to construct $\sqrt{K}^{-1}$. In what follows, we only focus on K^{-1}, but our construction can be applied to $\sqrt{K}^{-1}$ as well.

7.2 Texture-specific kernels

In [8] it was shown that the difference between the OSV model [42] and ROF model [44] could be understood as frequency weighting of the L^2 norm for the H^{-1} fidelity term of OSV. The frequency weighting of the square norm is proportional to $\frac{1}{\omega^2}$, which corresponds to the Δ^{-1} operator in the frequency domain, see Fig. 10 . The low frequencies are therefore highly penalized in the fidelity term, considerably reducing the eroding effect compared with ROF. This has proved to be an efficient tool for image denoising [42, 7]. In [8] it was suggested that other linear kernels could be used for adaptive frequency algorithms.

In this section we address the problem of designing a family of kernels for image decomposition. The operator K is a convolution operator, therefore K^{-1} in the Fourier domain is simply its inverse. Moreover, K^{-1} is also a convolution operator. We denote by H the associated filter, and in the rest of the section we focus on the designing of this filter.

In the $u+v$ decomposition model K penalizes frequencies that are not considered as part of the texture component. Therefore K^{-1} could be interpreted as the frequencies which should mainly be included in the texture part. A general and simple characterization of textures could be done using Gabor functions. These functions would typically describe the type of textures we would like to extract. Naturally, they apply as good candidates for K^{-1}. As already mentioned, the inverse kernel is actually the one needed in the numerical implementation. Thus our proposed design strategy is to use Gabor functions for constructing the inverse kernel. Notice that other design methods could be used. We use the function:

$$g(x) = \cos(2\pi\nu x)\,\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{55}$$

This gives the following values for the filter H:

$$h_k = \cos(2\pi\nu k)\,\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-k^2}{2\sigma^2}\right) \tag{56}$$

$\nu \in (0, 0.5]$ is the frequency of the texture. σ is related to the width of the band-pass around this frequency. A small σ in the spatial domain means a wide band-pass in the frequency domain. If we know the frequency of the texture we want to get, it is then interesting to use a large σ (which means a small band-pass in the frequency domain). Note that some restrictions apply for choosing σ, see Lemma 4. Actually, σ cannot

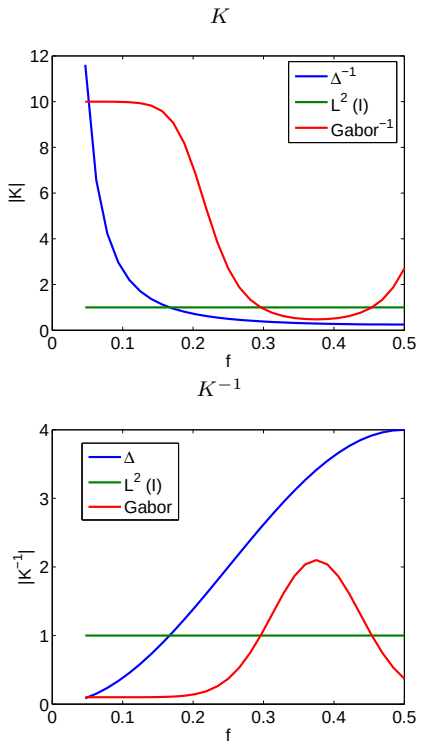


Figure 10: The kernel K and its inverse K^{-1} for the OSV, ROF and the proposed TV-Gabor model.

be very large, which may be interpreted as a from of an uncertainty principle.

(55) is a one dimension filter. There are many methods to then design a 2D filter. One possibility is to consider the product $g(x)g(y)$. We will analyze this possibility later. Another choice to construct our filter H is to use rotationally invariant Gabor wavelets as:

$$g(x,y) = \cos\left(2\pi\nu\sqrt{x^2+y^2}\right)\frac{1}{\sqrt{2\pi}\sigma^2}\exp\left(\frac{-x^2-y^2}{2\sigma^2}\right) \tag{57}$$

Such a choice will give better numerical results when the texture is known to be rotationally invariant.

Directions: Many textures are not rotationally invariant. It is therefore interesting to add this direction information in our filter H. To do so, we just need to use a $1D$ filter as (55), and then rotate it so that it fits the direction of the texture. A possible improvement is to use an ellipse (see [21] for instance).

7.3 1D and 2D filters

In this subsection, we propose a way to construct a 2D kernel K^{-1} (in fact of the associated filter H) out of a 1D filter:

$$H_x = \left(h_{\frac{d-1}{2}}, \ldots, h_1, h_0, h_1, \ldots, h_{\frac{d-1}{2}}\right) \tag{58}$$

where d is the dimension of the filter H_x, and h_k is given by (56). Since K^{-1} is symmetric, we also choose H_x to be symmetric. We then set $H = H_x * H_y$, where H stands for the filter associated to K^{-1}, $*$ denotes convolution, and $H_y = H_x^T$, where T stands for transpose.

Remark: In all this section, for the convolution, we consider periodic boundary conditions.

7.4 Eigenvalues

In this subsection, we compute the eigenvalues of K^{-1}, and give a sufficient condition so that they are positive.

The filter H associated with K^{-1} should define a linear symmetric positive operator. By construction, H defines a linear symmetric operator. But as we will see, we have to impose some conditions on the values h_k of the filter so that it is positive. We recall that a linear symmetric operator is positive if and only of its eigenvalues are positive (this can even be taken as a definition). To get the positivity for H, we are therefore lead to compute its associated eigenvalues (the ones of the associated linear mapping). Since we have constructed H out of two 1-D filters, we are in fact interested in the eigenvalues of these filters (since they will give us the eigenvalues of K^{-1}). Since K^{-1} is positive, we also impose the constraint that H_x is positive.

The filtering of an image of size $N \times M$ by H_x corresponds to a linear mapping from $\mathbb{R}^{NM}$ to $\mathbb{R}^{NM}$ (this is the reason why we speak of the eigenvalues of the filter H, which are in fact the eigenvalues of the corresponding linear mapping). Let us denote by A_x (resp A_y) the matrix of size $(NM)^2$ associated to H_x (resp H_y). An image I is a matrix $(I_{i,j})$, with $1 \le i \le N$ and $1 \le j \le M$. We rewrite it as a 1 Dimensional vector I_k, with $1 \le k \le NM$, using $I_k = I_{i,j}$ if $k = M(i-1)+j$.

Since A_x and A_y have a very particular form (they are both circulant matrices), we can compute the exact values of their eigenvalues, as stated by the following result:

Proposition 7 *The eigenvalues of A_x are:*

$$\left\{ h_0 + 2\sum_{k=1}^{\frac{d-1}{2}} h_k \cos\left(\frac{2\pi qk}{M}\right),\ 0 \le q \le \frac{M}{2} \right\} \tag{59}$$

and the ones of A_y are:

$$\left\{ h_0 + 2\sum_{k=1}^{\frac{d-1}{2}} h_k \cos\left(\frac{2\pi qk}{N}\right),\ 0 \le q \le \frac{N}{2} \right\} \tag{60}$$

Proof. The proof is just a consequence of the fact that A_x and A_y are circulant matrix. We refer the interested reader to [9] for the details. □

Now that we have computed the eigenvalues of A_x and A_y, we can get the ones of K^{-1}. Since A_x and A_y commute, the eigenvalues of K^{-1} are contained in the set:

$$\left\{ P_1(\omega_M^p) P_2(\omega_N^q),\ 0 \leq q \leq \frac{M}{2},\ 0 \leq q \leq \frac{N}{2} \right\} \quad (61)$$

Since the eigenvalues of A_x and A_y are positive, so are the ones of K^{-1}. If we denote by $\gamma^x_{\min}$ (resp γ^y_{min}) the smallest eigenvalue of A_x (resp A_y) and by $\gamma^x_{\max}$ (resp $\gamma^y_{\max}$) the largest eigenvalue of A_x (resp A_y), then, if γ is an eigenvalue of K^{-1}, we have:

$$\gamma^x_{min}\gamma^y_{min} \leq \gamma \leq \gamma^x_{\max}\gamma^y_{\max} \quad (62)$$

From this last point, we deduce in particular that

$$\|K^{-1}\|_{L^2} \leq \gamma^x_{\max}\gamma^y_{\max} \quad (63)$$

Lemma 3 *If we choose* $\tau \leq \frac{1}{8\gamma^x_{\max}\gamma^y_{\max}}$ *in algorithm (28), then the algorithm converges.*

Proof. This a direct consequence of (63) and of Theorem 1 (in Section 3). □

Unfortunately, the eigenvalues of K^{-1} can be negative. The next lemma gives a sufficient condition for the eigenvalues of K^{-1} to be positive.

Lemma 4 *If*

$$h_0 \geq 2 \sum_{k=1}^{\frac{d-1}{2}} |h_k| \quad (64)$$

then the eigenvalues of A_x, A_y *and* K^{-1} *are positive.*

Proof. This is a consequence of Proposition 7 and of (61). □

Notice that (64) is only a sufficient condition. The eigenvalues can still be positive in less restrictive cases, and can be computed explicitly for the designed kernel (see Proposition 7).

By using Lemma 4 and the explicit values of h_k given by (56), we can derive more explicit sufficient conditions about the positivity of the eigenvalues of K^{-1}. In particular, we show that if σ is small enough, then the eigenvalues of H are positive, see more details in [9].

7.5 Numerical results

We show some numerical results obtained with the new TV-Gabor model on Figures 11 to 16.

In Figure 11, the texture is a periodic signal of frequency $1/\pi \approx 0.32$. In this case we use a rotationally symmetric Gabor function of frequency 0.25 and $\sigma = 1$ (no directional knowledge is incorporated). As expected, the decomposition in this case is very good. In the next two examples we focus on the ability of the model to have directional selectivity of the textural part, a main feature that clearly distinguishes the TV-Gabor model from the previous ones. In case the textural directions are not known beforehand, we suggest to select them by the dominant peaks in the Fourier domain in medium and high frequencies. This can give basic but sufficient information for designing the kernel (choosing frequency and preferred direction). The Fourier transforms of the processed images are shown in Figures 14 and 15. In Figures 12 and 13 the original image is composed of two types of textures and a synthetic structural part. We would like to extract the periodic texture in the ellipses, and not the small squares on the top right. This type of selectivity is quite hard, but is achieved quite well, as seen in Fig. 12. Edges of the structural part are kept sharp, and clearly outperforms any linear kernel that would be designed to achieve a similar goal. Compared to $TV - L^2$ (Fig. 13, bottom) one observes that both textures are mostly in the v part. Also there is some more erosion of the structure (seen in the brighter triangle in the v component) and some "left-overs" of the ellipses-texture in the u part. The comparison was made such that both v parts of TV-Gabor and $TV - L^2$ have the same L^2 norm.

In Figures 15 and 16, we show another example of directional decomposition of part of a Dollar note image. In this case, we use the directional TV-Gabor model in the y direction to capture the forehead textures. For comparison, we also display the result with the standard $TV - L^2$ model. As the textures are quite fine with low contrast, we show in Fig. 16, bottom, a contrast enhanced version of v, by multiplying the v part by 4. Again here, both v components have the same L^2 norm. One clearly sees the high directional selectivity of the TV-Gabor model on the left, versus the non-selectiveness of $TV - L^2$.

8 Conclusion

In this paper, we have studied the problem of image decomposition. Given an original image f, we split it into two components u and v, u containing the geometric information and v the texture information. Our modeling is focused on TV regularization approaches: we minimize a functional with two terms, the first one is based on the total variation semi-norm and the second

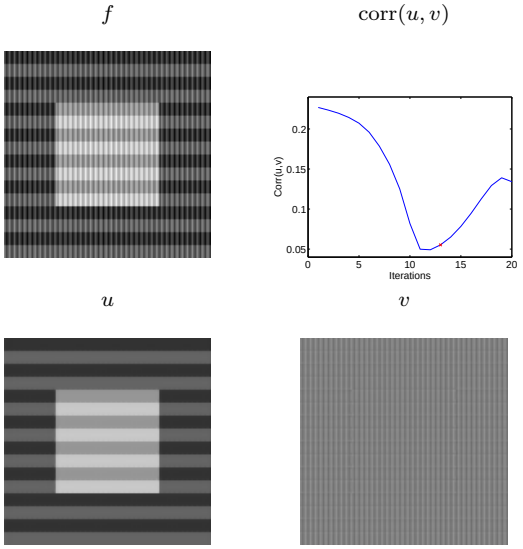


Figure 11: Decomposition of a simple image by TV-Gabor.

one on a different norm adapted to the texture component of the image. We have considered four different decomposition models: $TV - L^2$, $TV - G$, $TV - L^1$ and TV-Gabor.

An interesting conclusion of this study is a form of "recipe" we can derive to carry out image decomposition:

1. If the texture part is known to be very structured, then the $TV - L^1$ approach seems to be the best choice.

2. In the case of directional texture or if an estimation of the frequency of the texture is known, and if the texture is rather smooth, then the TV-Gabor model is the more appropriate approach.

3. In a general case, when no *a-priori* knowledge of the texture is at hand, we advocate the $TV - L^2$ approach, or its improvement with the $TV - G$ regularization.

This provides us with a sort of image decomposition toolbox for a wide class of synthetic and natural images.

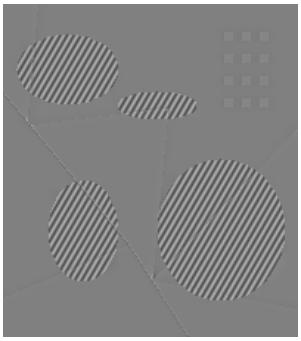
Apart from the broad perspective and our suggestions when each model should be used, the main contributions of our paper are related to the three following issues:

1. First, we show that the correlation graph between u and v is an efficient tool to select the splitting parameter. We have applied this method to the four

Figure 12: u component of the decomposition of a synthetic image with textures of specific frequency and orientation by TV-Gabor and $TV - L^2$. The TV-Gabor can be more selective and reduce the inclusion in v of undesired textures / small-structures like the small blocks on the top right. Also erosion of large structures is reduced (more apparent in the brighter triangle).

models. As far as we know, this is the first attempt

TV-Gabor, v



TV-L^2, v

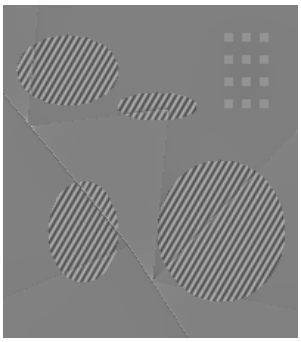


Figure 13: v component of the decomposition of a synthetic image with textures of specific frequency and orientation by TV-Gabor (top) and $TV - L^2$ (bottom). See the caption of Fig. 12.

to tune the decomposition parameter of such models other than by trial and error (the problem had been considered before only in the denoising case).

2. Second, we propose new and fast algorithms to solve the TV-L^1 minimization problem using projection and thresholding techniques. We have carried out the complete mathematical study of these algorithms, and shown their efficiency with some numerical examples.

3. Third, we introduce a new TV-Gabor model which leads us to adaptive frequency and directional image decomposition. In the case when we have some additional information about the texture, then we can take advantage of it by incorporating this information in the functional. We have designed and studied the corresponding filters, and we have illustrated this new approach with numerical examples.

Fourier transform of f $\quad$ $\mathrm{corr}(u, v)$

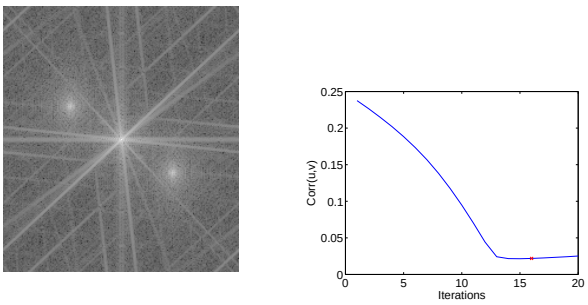


Figure 14: Fourier transform and the correlation of TV-Gabor of the synthetic image in Fig. 12.

In this paper we presented a way to design simple texture-specific filters based on Gabor functions. Other, more sophisticated methods could be incorporated to this framework, such as ones based on wavelets [45]. In future works we intend to explore these issues. Notice that a straightforward extension of the new TV-Gabor model to multiple selected directions, is to use the linearity of the Hilbert fitting term and simply add several directional kernels.

A natural generalization for the $u + v$ decomposition is to consider a multi-scale approach, as done in [49, 26, 25, 41, 29]. This also relates to the parameter selection problem, where better and more accurate mechanisms could be used instead of the correlation criterion. A more detailed version of this work, with some more theoretical results and proofs can be found in our report [9].

Acknowledgements: Jean-François Aujol and Tony Chan acknowledge supports by grants from the NSF under contracts DMS-9973341, ACI-0072112, INT-0072863, the ONR under contract N00014-03-1-0888, the NIH under contract P20 MH65166, and the NIH Roadmap Initiative for Bioinformatics and Computational Biology U54 RR021813 funded by the NCRR, NCBC, and NIGMS. Guy Gilboa acknowledges support by the following grants: NIH U54 RR021813, NSF IIS-0326388 (Prime award), NYU F5552-01. Stanley Osher acknowledges support by the following grants: NIH U54 RR021813, NSF IIS-0326388 (Prime award), NYU F5552-01, NSF DMS-0312222, and NSF ACI-0321917.

The authors would like to thank Wotao Yin for fruitful discussions, as well as for the numerical results obtained with the algorithm of [51] presented on Figure 9.

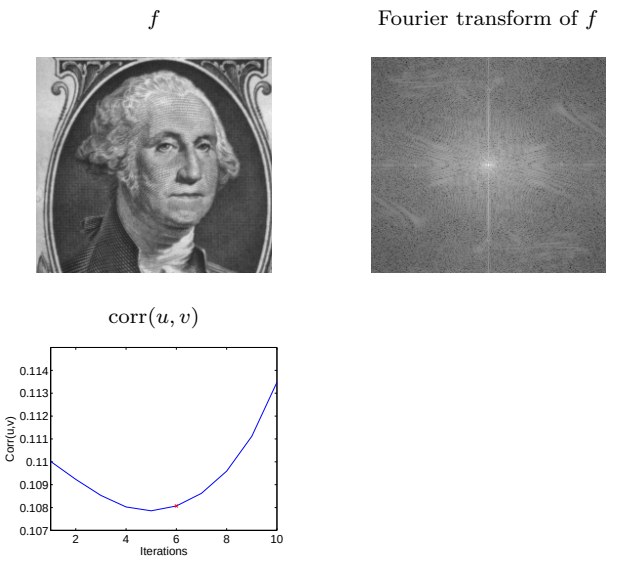


Figure 15: Decomposition of a Dollar note image by directional TV-Gabor in the y direction to capture the forehead textures.

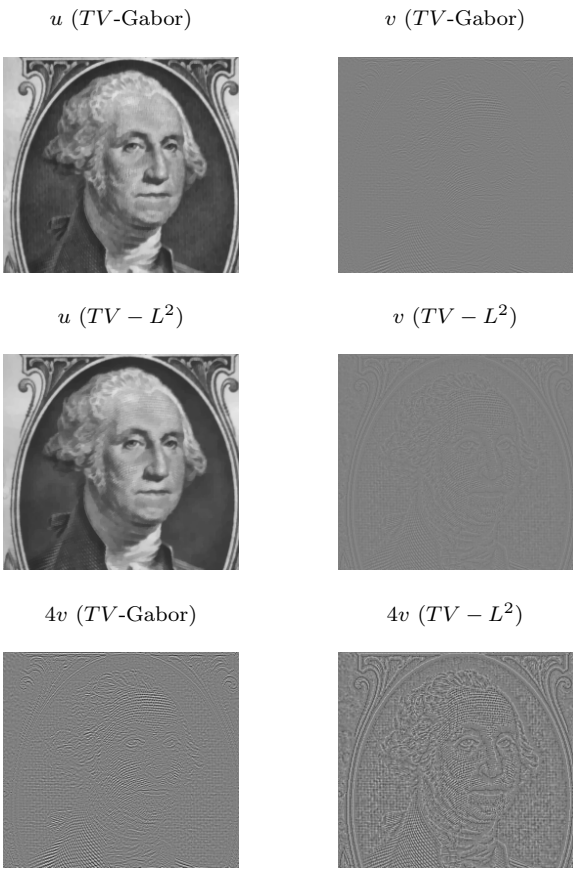


Figure 16: Decomposition of a Dollar note image by TV-Gabor in the y direction and by TV-L^2.

A Proofs for the $TV - L^1$ algorithm

In this appendix, we give the proofs of the Mathematical results stated in Section 6 for the new fast $TV - L^1$ algorithm.

A.1 Existence and uniqueness of a solution

We give here the proof of Theorem 2 stated in Section 6. We first recall the theorem:

Theorem 2 *Problem (44) admits a unique solution* $(\tilde{u}, \tilde{v})$ *in* $(X \times X)$.

To prove Theorem 2, we will use the following lemma:

Lemma 5 *Let us assume that* $f \neq 0$ *(i.e. that there exists* (i, j) *such that* $f_{i,j} \neq 0$*). If* $g \in X$*, then* $(0, g)$ *is not a minimizer of problem (44).*

Proof.(*Lemma 5*): By contradiction, let us assume that there exists $g \in X$ such that $(0, g)$ is a minimizer of problem (44), then in particular we have:

$$M(0, g) = \inf_v \left(\frac{1}{2\alpha} \|f - v\|_{L^2}^2 + \lambda \|v\|_{L^1} \right) \qquad (65)$$

From Proposition 4, we get that $g = ST(f, \alpha\lambda)$. Since $f \geq 0$ (in our case, we even have $0 \leq f \leq 255$), we deduce that $g \geq 0$.

- Let us first assume that there exists (i, j) such that $g_{i,j} > 0$. Let us define $\epsilon = \min(g_{i,j} \text{ such that } g_{i,j} > 0)$. We have $M(\epsilon, g - \epsilon) = M(0, g) - N^2\epsilon$ (we recall that f is of size $N \times N$), which contradicts the fact that $(0, g)$ is a minimizer of problem (44).
- Let us now assume that $g_{i,j} = 0$ for all $(i, j) \in N^2$. We define $\epsilon_1 = \frac{\int f}{N^2}$ (we know that $\int f > 0$). We have $M(\epsilon_2, 0) = \frac{1}{2\alpha}\|f - \epsilon_2\|_{L^2}^2$. But $\|f - \epsilon_2\|_{L^2}^2 = \|f\|_{L^2}^2 + N^2\epsilon_2^2 - 2\epsilon_2 \int f < \|f\|_{L^2}^2$. Therefore, we get $M(\epsilon_2, 0) < \frac{1}{2\alpha}\|f\|_{L^2}^2 = M(0, 0)$, which contradicts the fact that $(0, g)$ is a minimizer of problem (44).

□

Proof.(*Theorem 2*): The existence of a solution for problem (44) is standard. It is a straightforward con-

sequence of the fact that M the functional to minimize is convex and coercive.

Let us now show the uniqueness. In the case when $f = 0$, then it is clear that $(0,0)$ is the unique minimizer of problem (44). Let us therefore assume that $f \neq 0$ (i.e. that there exists (i,j) such that $f_{i,j} \neq 0$). By contradiction, let us assume that there exist two solutions for problem (44), (u_1, v_1) and (u_2, v_2). We denote by $m = M(u_1, v_1) = M(u_2, v_2)$. If $t \in (0,1)$, then we get:

$$\begin{aligned} &M(tu_1 + (1-t)u_2, tv_1 + (1-t)v_2) = \\ &\|tu_1 + (1-t)u_2\|_{\dot{B}^1_{1,1}} + \lambda\|tv_1 + (1-t)v_2\|_{L^1} \\ &+\tfrac{1}{2\alpha}\|t(f - u_1 - v_1) + (1-t)(f - u_2 - v_2)\|^2_{L^2} \end{aligned} \tag{66}$$

But by convexity, we have:

$$\|tu_1 + (1-t)u_2\|_{\dot{B}^1_{1,1}} \leq t\|u_1\|_{\dot{B}^1_{1,1}} + (1-t)\|u_2\|_{\dot{B}^1_{1,1}} \tag{67}$$

and

$$\|tv_1 + (1-t)v_2\|_{L^1} \leq t\|u_1\|_{L^1} + (1-t)\|u_2\|_{L^1} \tag{68}$$

as well as

$$\begin{aligned} &\|t(f - u_1 - v_1) + (1-t)(f - u_2 - v_2)\|^2_{L^2} \\ &\leq t\|f - u_1 - v_1\|^2_{L^2} + (1-t)\|f - u_2 - v_2\|^2_{L^2} \end{aligned} \tag{69}$$

From (66)-(67)-(68)-(69), we deduce that (since $M(u_1, v_1) = M(u_2, v_2) = m$):

$$M(tu_1 + (1-t)u_2, tv_1 + (1-t)v_2) \leq m \tag{70}$$

and (70) is an equality if and only if (67)-(68)-(69) are equalities. But by definition, we have $M(tu_1 + (1-t)u_2, tv_1 + (1-t)v_2) \geq m$. Therefore (70) must be an equality, as well as (67)-(68)-(69).

The function in (69) is strictly convex. Therefore (69) is an equality if and only if $f - u_1 - v_1 = f - u_2 - v_2$, i.e. if and only if

$$u_1 - u_2 = v_2 - v_1 \tag{71}$$

(67) is an equality if and only if there exists $w_u \in X\backslash\{0\}$ and $(a_u, b_u) \in \mathbb{R}^2_+$ such that $u_1 = a_u w_u$ and $u_2 = b_u w_u$. (68) is an equality if and only if there exists $w_v \in X\backslash\{0\}$ and $(a_v, b_v) \in \mathbb{R}^2_+$ such that $v_1 = a_v w_v$ and $v_2 = b_v w_v$. Using w_u and w_v, then (71) becomes $(a_u - b_u)w_u = (a_v - b_v)w_v$. Since we assume that $(u_1, v_1) \neq (u_2, v_2)$, this implies that we cannot have simultaneously $a_u = b_u$ and $a_v = b_v$. We thus get that w_u and w_v are proportional.

We therefore deduce that there exists $w \in X\backslash\{0\}$ and $(a, b, c) \in \mathbb{R}^4$ such that $u_1 = aw$, $u_2 = bw$, $v_1 = cw$ and $v_2 = (a - b + c)w$. Moreover, we have $a - b \neq 0$. Let us remark that: $M(tu_1 + (1-t)u_2, tv_1 + (1-t)v_2) = M(u_2 + t(u_1 - u_2), v_2 + t(v_1 - v_2)) = M(u_1 + (t-1)(u_1 - u_2), v_1 + (t-1)(v_1 - v_2))$. We recall that $t \in (0,1)$. We assume that $a \neq 0$ and $b \neq 0$ (in the case when $a = 0$ or $b = 0$, then we get a contradiction thanks to Lemma 5). We impose $0 \leq t < \min\left(1, \frac{|a|}{|a-b|}, \frac{|b|}{|a-b|}\right)$:

$$\begin{aligned} 0 &\leq M(u_2 + t(u_1 - u_2), v_2 + t(v_1 - v_2)) - M(u_2, v_2) \\ &= \|aw + t(a-b)w\|_{\dot{B}^1_{1,1}} - \|aw\|_{\dot{B}^1_{1,1}} \\ &\quad +\lambda\|bw - t(a-b)w\|_{L^1} - \lambda\|bw\|_{L^1} \\ &= t|a-b|\left(\|w\|_{\dot{B}^1_{1,1}} - \lambda\|w\|_{L^1}\right) \end{aligned}$$

We therefore deduce that $\|w\|_{\dot{B}^1_{1,1}} - \lambda\|w\|_{L^1} \geq 0$.

By using the fact that $0 \leq M(u_1 + (t-1)(u_1 - u_2), v_1 + (t-1)(v_1 - v_2)) - M(u_1, v_1)$, we get exactly as before that $\|w\|_{\dot{B}^1_{1,1}} - \lambda\|w\|_{L^1} \leq 0$. We therefore deduce that:

$$\|w\|_{\dot{B}^1_{1,1}} = \lambda\|w\|_{L^1} \tag{72}$$

And (72) also holds with u_1, u_2, v_1 and v_2. In particular, this implies that $(0, u_1 + v_1)$ is a minimizer of problem (44). Since we assume that $f \neq 0$ (i.e that there exists (i,j) such that $f_{i,j} \neq 0$), we get a contradiction thanks to Lemma 5. □

A.2 Convergence of the $TV - L^1$ algorithm

We give here the proofs of Proposition 5 and 6 stated in Section 6.

Proof. (*Proposition 5):*

The proof uses the same ideas as the ones of Proposition 3.4 in [5], but we put it here for the sake of completeness.

We first remark that, as we solve successive minimization problems, we have:

$$M(u_n, v_n) \geq M(u_n, v_{n+1}) \geq M(u_{n+1}, v_{n+1}) \tag{73}$$

In particular, the sequence $M(u_n, v_n)$ is nonincreasing. As it is bounded from below by 0, it thus converges in $\mathbb{R}$. We denote by m its limit. We want to show that

$$m = \inf_{(u,v)\in X\times X} M(u, v) \tag{74}$$

As M is coercive and as the sequence $M(u_n, v_n)$ converges, we deduce that the sequence (u_n, v_n) is bounded in $X \times X$. We can thus extract a subsequence (u_{n_k}, v_{n_k}) which converges to $(\hat{u}, \hat{v})$ as $n_k \to +\infty$, with $(\hat{u}, \hat{v}) \in X \times X$. Moreover, we have, for all $n_k \in \mathcal{N}$ and all v in X:

$$M(u_{n_k}, v_{n_k+1}) \leq M(u_{n_k}, v) \tag{75}$$

and for all $n_k \in \mathcal{N}$ and all u in X:

$$M(u_{n_k}, v_{n_k}) \leq M(u, v_{n_k}) \tag{76}$$

Let us denote by $\tilde{v}$ a cluster point of (v_{n_k+1}). Considering (73), we get (since M is continuous on $X \times X$):

$$m = M(\hat{u}, \hat{v}) = M(\hat{u}, \tilde{v}) \tag{77}$$

By passing to the limit in (48), we get: $\tilde{v} = ST(f - \hat{u}, \alpha\lambda)$, i.e. $\tilde{v}$ is the solution of $\inf_v \left(\frac{1}{2\alpha}\|f - \hat{u} - v\|_2^2 + \lambda\|v\|_{L^1}\right)$. But from (77), we know that: $\frac{1}{2\alpha}\|f - \hat{u} - \tilde{v}\|_2^2 + \lambda\|\tilde{v}\|_{L^1} = \frac{1}{2\alpha}\|f - \hat{u} - \hat{v}\|_2^2 + \lambda\|\hat{v}\|_{L^1}$. By uniqueness of the solution, we conclude that $\tilde{v} = \hat{v}$. Hence $v_{n_k+1} \to \hat{v}$. By passing to the limit in (75) (M is continuous on $X \times X$), we therefore have for all v:

$$M(\hat{u}, \hat{v}) \leq M(\hat{u}, v) \tag{78}$$

And by passing to the limit in (76), for all u:

$$M(\hat{u}, \hat{v}) \leq M(u, \hat{v}) \tag{79}$$

(78) and (79) can respectively be rewritten:

$$M(\hat{u}, \hat{v}) = \inf_{v \in X} M(\hat{u}, v) \tag{80}$$

$$M(\hat{u}, \hat{v}) = \inf_{u \in X} M(u, \hat{v}) \tag{81}$$

But, from the definition of $M(u, v)$, (81) is equivalent to (see [22]):

$$0 \in -f + \hat{u} + \hat{v} + \alpha\partial J_B(\hat{u}) \tag{82}$$

and (80) to:

$$0 \in -f + \hat{u} + \hat{v} + \alpha\lambda\partial J_{L1}(\hat{v}) \tag{83}$$

where the functions J_B is defined by $J_B(u) = \|u\|_{\dot{B}^1_{1,1}}$ and J_{L1} by $J_{L1}(v) = \|v\|_{L^1}$. The subdifferential of M at $(\hat{u}, \hat{v})$ is given by:

$$\partial M(\hat{u}, \hat{v}) = \frac{1}{\alpha}\begin{pmatrix} -f + \hat{u} + \hat{v} + \alpha\partial J_B(\hat{u}) \\ -f + \hat{u} + \hat{v} + \alpha\lambda\partial J_{L1}(\hat{v}) \end{pmatrix} \tag{84}$$

And thus, according to (82) and (83), we have:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \partial M(\hat{u}, \hat{v}) \tag{85}$$

which is equivalent to: $M(\hat{u}, \hat{v}) = \inf_{(u,v)\in X^2} M(u, v) = m$. Hence the whole sequence $M(u_n, v_n)$ converges towards m the unique minimum of M on $X \times X$. We deduce that the sequence (u_n, v_n) converges to $(\hat{u}, \hat{v})$, the minimizer of M, when n tends to $+\infty$.

□

Proof. (*Proposition 6*):

The proof is very similar to the one of Proposition 3.8 in [5].

The existence and uniqueness of $(u_{\alpha_n}, v_{\alpha_n})$ is given by Theorem 2. Since $(u_{\alpha_n}, v_{\alpha_n})$ is the solution of problem (44), we have

$$M(u_{\alpha_n}, v_{\alpha_n}) \leq M(f, 0) \tag{86}$$

From this, we get that $(u_{\alpha_n}, v_{\alpha_n})$ is bounded. Then, up to an extraction, there exists $(u_0, v_0) \in X \times X$ such that $(u_{\alpha_n}, v_{\alpha_n})$ converges to (u_0, v_0). From (86), we get that $\|f - u_{\alpha_n} - v_{\alpha_n}\|_2^2 \leq 2\alpha\|f\|_{\dot{B}^1_{1,1}}$. By passing to the limit $n \to +\infty$, we get: $\|f - u_0 - v_0\|_2 = 0$, i.e. $v_0 = f - u_0$.

To conclude the proof of the proposition, there remains to show that $(u_0, f - u_0)$ is a solution of problem (52). Let $u \in X$. We have:

$$\begin{aligned} & \|u\|_{\dot{B}^1_{1,1}} + \lambda\|f - u\|_{L^1} + \tfrac{1}{2\alpha_n}\underbrace{\|f - u - (f - u)\|^2}_{=0} \\ \geq\ & \|u_{\alpha_n}\|_{\dot{B}^1_{1,1}} + \lambda\|v_{\alpha_n}\|_{L^1} + \tfrac{1}{2\alpha_n}\|f - u_{\lambda_n} - v_{\lambda_n}\|^2 \\ \geq\ & \underbrace{\|u_{\alpha_n})\|_{\dot{B}^1_{1,1}} + \lambda\|v_{\alpha_n}\|_{L^1}}_{\to \|u_0\|_{\dot{B}^1_{1,1}} + \lambda\|f - u_0\|_{L^1}} \end{aligned}$$

□

References

[1] R. Adams. *Sobolev Spaces*. Pure and applied Mathematics. Academic Press, Inc, 1975.

[2] S. Aliney. A property of the minimum vectors of a regularizing functional defined by means of the absolute norm. *IEEE Transactions on Signal Processing*, 45(4):913–917, 1997.

[3] G. Aubert and J.F. Aujol. Modeling very oscillating signals. Application to image processing. *Applied Mathematics and Optimization*, 51(2):163–182, March/April 2005.

[4] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2002.

[5] J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1):71–88, January 2005.

[6] J.F. Aujol, G. Aubert, L. Blanc-Féraud, and Antonin Chambolle. Decomposing an image: Application to SAR images. In *Scale-Space '03*, volume 2695 of *Lecture Notes in Computer Science*, 2003.

[7] J.F. Aujol and A. Chambolle. Dual norms and image decomposition models. *International Journal on Computer Vision*, 63(1):85–104, June 2005.

[8] J.F. Aujol and G. Gilboa. Implementation and parameter selection for BV-Hilbert space regularizations, 2004. UCLA CAM Report 04-66.

[9] J.F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition – modeling, algorithms, and parameter selection, 2005. UCLA CAM Report 05-10, ftp://ftp.math.ucla.edu/pub/camreport/cam05-10.pdf.

[10] J.F. Aujol and S.H. Kang. Color image decomposition and restoration, 2004. UCLA CAM Report 04-73, to appear in Journal of Visual Communication and Image Representation.

[11] J. Bect, G. Aubert, L. Blanc-Féraud, and A. Chambolle. A l^1 unified variational framwork for image restoration. In *ECCV 2004*, May 2004.

[12] A. Chambolle. An algorithm for total variation minimization and applications. *JMIV*, 20:89–97, 2004.

[13] A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(3):167–188, 1997.

[14] A. Chambolle, R.A. De Vore, N. Lee, and B.J. Lucier. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transcations on Image Processing*, 7(3):319–335, March 1998.

[15] T. Chan and S. Esedoglu. Aspects of total variation regularized L^1 function approximation, 2004. CAM report 04-07, to appear in SIAM Journal on Applied Mathematics.

[16] P. L. Combettes and V. R. Wajs. Theoretical analysis of some regularized image denoising methods. In *ICIP 04*, volume 1, pages 969–972, 2004.

[17] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Journal on Multiscale Modeling and Simulation*, 2005. *in press.*

[18] P.L. Combettes and J. Luo. An adaptative level set method for nondifferentiable constrained image recovery. *IEEE Transactions on Image Processing*, 11(11):1295–1304, 2002.

[19] I. Daubechies and G. Teschke. Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising, 2004. submitted.

[20] D.L. Donoho and M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, December 1995.

[21] D. Dunn and W.E. Higgins. Optimal Gabor filters for texture segmentation. *IEEE Transactions on Image Processing*, 4(7):947–964, July 1995.

[22] I. Ekeland and R. Temam. *Analyse convexe et problmes variationnels*. Etudes Mathématiques. Dunod, 1974.

[23] D. Gabor. Theory of communication. *J. Inst. of Electrical Engineering*, 93(3):429–457, 1946.

[24] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Estimation of optimal PDE-based denoising in the SNR sense. submitted.

[25] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Variational denoising of partly-textured images by spatially varying constraints. submitted.

[26] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Texture preserving variational denoising using an adaptive fidelity term. In *Proc. VLSM 2003, Nice, France*, pages 137–144, 2003.

[27] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Estimation of optimal PDE-based denoising in the SNR sense, 2004. CCIT report No. 499, Technion, August, see http://www.math.ucla.edu/∼gilboa/.

[28] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Estimation of the optimal variational parameter via SNR analysis. In *Scale-Space '05*, volume 3459 of *Lecture Notes in Computer Science*, pages 230–241, April 2005.

[29] C. Groetsch and O. Scherzer. Inverse scale space theory for inverse problems. In *Scale-Space '01*, volume 2106 of *Lecture Notes in Computer Science*, pages 317–325, 2001.

[30] M. Hintermuller and K. Kunisch. Total bounded variation regularization as bilaterally constrained optimization problem. *SIAM Journal on Applied Mathematics*, 64(4):1311–1333, 2004.

[31] A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[32] T. Le and L. Vese. Image decomposition using total variation and div(BMO), 2004. UCLA CAM Report 04-36.

[33] F. Malgouyres. Mathematical analysis of a model which combines total variation and wavelet for image restoration. *Journal of information processes*, 2(1):1–10, 2002.

[34] F. Malgouyres. Minimizing the total variation under a general convex constraint for image restoration. *IEEE transactions on image processing*, 11(12):1450–1456, December 2002.

[35] S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[36] S.G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[37] Yves Meyer. Oscillating patterns in image processing and in some nonlinear evolution equations, March 2001. The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures.

[38] V. A. Morosov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417, 1966.

[39] P. Mrázek and M. Navara. Selection of optimal stopping time for nonlinear diffusion filtering. *IJCV*, 52(2/3):189–203, 2003.

[40] M. Nikolova. A variational approach to remove outliers and impulse noise. *JMIV*, 20(1-2):99–120, 2004.

[41] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation based image restoration, 2004. CAM report 04-13, to appear in SIAM Journal on Multiscale Modeling and Simulation.

[42] S.J. Osher, A. Sole, and L.A. Vese. Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, 1(3):349–370, 2003.

[43] M. Porat and Y.Y. Zeevi. The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):452–468, July 1988.

[44] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[45] J.L. Starck, M. Elad, and D.L. Donoho. Image decomposition: separation of texture from piecewise smooth content, 2003. To appear in IEEE Transactions on Image Processing.

[46] G. Steidl, S. Didas, and J. Neumann. Relations between higher order TV regularization and support vector regression. In *Scale-Space and PDE methods in computer vision, R. Kimmel, N. Sochen, J. Weickert Eds*, volume LNCS 3459, pages 515–527, 2005.

[47] G. Steidl, J. Weickert, T. Brox, P. Mrazek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides. *SIAM J. Numer. Anal.*, 42:686–658, 2004.

[48] D. Strong, J.F. Aujol, and T.F. Chan. Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization, January 2005. UCLA CAM Report 05-02.

[49] E. Tadmor, S. Nezzar, and L. Vese. A multiscale image representation using hierarchical (BV, L^2) decompositions. *SIAM Journal on Multiscale Modeling and Simulation*, 2(4):554–579, 2004.

[50] L.A. Vese and S.J. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19:553–572, 2003.

[51] W. Yin, D. Goldfarb, and S. Osher. Total variation based image cartoon-texture decomposition, 2005. Columbia University CORC Report TR-2005-01, UCLA CAM Report 05-27.

[52] M. Zibulski and Y. Y. Zeevi. Analysis of multiwindow Gabor-type schemes by frame methods. *J. Appl. Comp. Harmon. Anal.*, 4(2):188–221, 1997.