Total variation based image denoising and restoration

Vicent Caselles*

Abstract. This paper is devoted to the total variation (TV) based approach to image denoising and restoration. The restored image minimizes total variation in the class of images which satisfy the contraints given by the image acquisition model. We compute some explicit solutions of the denoising model which explain some of the features observed in numerical experiments. We also comment on some alternatives recently proposed by Y. Meyer which lead to $u + v$ image decompositions. Finally we propose a total variation approach to image restoration, i.e., deconvolution and denoising, in which the image acquisition model is incorporated as a set of local constraints.

Mathematics Subject Classification (2000). Primary 68U10; Secondary 65K10, 65J20, 94A08.

Keywords. Image restoration, total variation, variational methods.

1. Introduction

We assume that the image acquisition system may be modelled by the following image formation model

$$z = h * u + n, \tag{1}$$

where $u\colon \mathbb{R}^2 \to \mathbb{R}$ denotes the ideal undistorted image, $h\colon \mathbb{R}^2 \to \mathbb{R}$ is a blurring kernel, z is the observed image which is represented as a function $z\colon \mathbb{R}^2 \to \mathbb{R}$, and n is an additive Gaussian white noise with zero mean and standard deviation σ.

Let us denote by Ω the interval $(0, N]^2$. As in most of works, in order to simplify this problem, we shall assume that the functions h and u are periodic of period N in each direction. That amounts to neglecting some boundary effects. Therefore, we shall assume that h, u are functions defined in Ω and, to fix ideas, we assume that $h, u \in L^2(\Omega)$. Our problem is to recover as much as possible of u, from our knowledge of the blurring kernel h, the statistics of the noise n, and the observed image z.

The problem of recovering u from z is ill-posed due to the ill-conditioning of the operator $Hu = h * u$. Several methods have been proposed to recover u. Most of them can be classified as regularization methods which may take into account statistical properties (Wiener filters), information theoretic properties ([19]), a priori

*The author acknowledges partial support by the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya and by PNPGC project, reference BFM2003-02125.

Proceedings of the International Congress
of Mathematicians, Madrid, Spain, 2006

geometric models ([30]) or the functional analytic behavior of the image given in terms of its wavelet coefficients ([20]).

The typical strategy to solve this ill-conditioning is regularization. Probably one of the first examples of regularization method [31] consists in choosing between all possible solutions of (1) the one which minimized the Sobolev (semi) norm of u

$$\int_\Omega |Du|^2\,dx. \tag{2}$$

Usually, the only information we know about the noise is statistical and limited to an estimate of its mean and its variance. In that case, the model equation (1) is incorporated as a set of constraints for (2): a first constraint corresponding to the assumption that the noise has zero mean, and a second one translating the fact that σ is an upper bound of the standard deviation of n.

This formulation was an important step, but the results were not satisfactory, mainly due to the unability of the previous functional to resolve discontinuities (edges) and oscillatory textured patterns. The smoothness required by the Dirichlet integral (2) is too restrictive and information corresponding to high frequencies of z is attenuated by it. Indeed, functions in $W^{1,2}(\Omega)$ (i.e., functions $u \in L^2(\Omega)$ such that $Du \in L^2(\Omega)$) cannot have discontinuities along rectifiable curves. These observations motivated the introduction of total variation in image restoration problems by L. Rudin, S. Osher and E. Fatemi in their work [30]. The a priori hypothesis is that functions of bounded variation (the BV model) ([5]) are a reasonable functional model for many problems in image processing, in particular, for restoration problems ([30]). Typically, functions of bounded variation have discontinuities along rectifiable curves, being continuous in some sense (in the measure theoretic sense) away from discontinuities. The discontinuities could be identified with edges. The ability of total variation regularization to recover edges is one of the main features which advocates for the use of this model but its ability to describe textures is less clear, even if some textures can be recovered, up to a certain scale of oscillation.

On the basis of the BV model, Rudin–Osher–Fatemi [30] proposed to solve the following constrained minimization problem

$$\begin{aligned} &\text{Minimize} \quad \int_\Omega |Du| \\ &\text{subject to} \quad \int_\Omega |h*u(x) - z(x)|^2\,dx \le \sigma^2|\Omega|. \end{aligned} \tag{3}$$

Notice that the image acquisition model (1) is only incorporated through a global constraint. Notice also that, assuming that $h*1 = 1$ (energy preservation), the constraint that $\int_\Omega h*u\,dx = \int_\Omega z(x)$ is automatically satisfied by its minima [17]. In practice, the above problem is solved via the following unconstrained minimization problem

$$\text{Minimize} \quad \int_\Omega |Du| + \frac{\lambda}{2}\int_\Omega |h*u - z|^2\,dx \tag{4}$$

where the parameter λ is positive. Recall that we may interpret λ^{-1} as a penalization parameter which controls the trade-off between the goodness of fit of the constraint and the smoothness term given by the total variation. In this formulation, a methodology is required for a correct choice of λ. The connections between (3) and (4) were studied by A. Chambolle and P. L. Lions in [17] where they proved that both problems are equivalent for some positive value of the Lagrange multiplier λ.

A particular and important case contained in the above formulation is the denoising problem which corresponds to the case where $h = \delta$, so that equation (1) is written as $z = u + n$ where n is an additive Gaussian white noise of zero mean and variance σ^2. In this case, the unconstrained variational formulation (5) with $h = \delta$ is

$$\text{Minimize} \quad \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega |u - z|^2 \, dx, \tag{5}$$

and it has been the object of much theoretical and numerical research (see [7] for a survey). Even if this model represented a theoretical and practical progress in the denoising problem due to the introduction of BV functions as image models, the experimental analysis readily showed its main drawbacks. Between them, let us mention the staircasing effect (when denoising a smooth ramp plus noise, the staircase is an admissible result), the pixelization of the image at smooth regions and the loose of fine textured regions, to mention some of them. This can be summarized with the simple observation that the residuals $z - u$, where u represents the solution of (5), do not look like noise. The theoretical analysis of the behavior of solutions of (5) has been the objects of several works [3], [12], [13], [27], [26] and will be developed in Section 2 by exhibiting explicit solutions for specially constructed functions z.

In spite of this, a second life in the interest of total variation based regularization was initiated after the proposal of $u + v$ models by Y. Meyer in [26]. The solution u of (5) permits to obtain a decomposition of the data z as a sum of two components $u + v$ where v is supposed to contain the noise and textured parts of the image z, while u contains the geometric sketch of the image z. As Meyer observed, the L^2 norm of the residual $v := z - u$ in (5) is not the right one to obtain a decomposition of z in terms of geometry plus texture and he proposed to measure the size of the textured part v in terms of a dual BV norm showing that some models of texture have a small dual BV norm: this will be the object of Section 3.

The restoration problem (which corresponds to the case of nontrivial kernel h) has also been the object of much interest due to its applications in many contexts, like satellite, astronomical or video images, to mention a few of them. In Section 4 we shall discuss a total variation based approach to the restoration model in which the image acquisition model is incorporated as a set of local constraints. Indeed, when incorporating (1) as a constraint in (3) we loose the local character of (1) and the restored image does not look satisfactory in textured and smooth regions at the same time. Thus, we propose to incorporate (1) by ensuring that the residuals $z - h * u$ have a variance bounded by σ^2 in a sufficiently large region around each pixel (the sampling process is incorporated in the model), the size of the region has to be sufficient in

order to estimate the variance of the noise. This gives a constrained formulation of the problem with as many Lagrange multipliers as pixels, and a solution is computed using Uzawa's method. Finally, in Section 5 we display some experiments on restoration of satellite images which illustrate the results that can be obtained with this method.

2. Explicit solutions of TV based denoising

The constrained formulation of the total variation denoising is given by (3) with $h = \delta$. Its unconstrained formulation is given by (5) where $\lambda > 0$ is a penalization parameter. Both problems are equivalent for a certain value of λ [17]. Our purpose in this section is to exhibit some qualitative features of total variation denoising by constructing explicit solutions of (5). Those features are well known at the experimental level, and the results give a theoretical justification of these observations. Our solutions will exhibit the possibility to resolve discontinuities, but also the loss of contrast, and the regularization of corners (thus, the image is loosing structure). The staircasing effect was explained in [27].

The construction of explicit solutions of (5) is related to the computation of solutions of the eigenvalue problem for the 1-Laplacian operator.

$$-\operatorname{div}\left(\frac{Du}{|Du|}\right) = u. \tag{6}$$

We denote by $\mathrm{BV}(\mathbb{R}^N)$ the space of functions of bounded variation in $\mathbb{R}^N$. For definitions concerning bounded variation functions we refer to [5]. The solution of (6) is understood in the following sense ([6], [7], [13]).

Definition 2.1. We say that a function $u \in L^2(\mathbb{R}^N) \cap \mathrm{BV}(\mathbb{R}^N)$ is a solution of (6) in $\mathbb{R}^N$ if there exits a vector field $\xi \in L^\infty(\mathbb{R}^N; \mathbb{R}^N)$ with $\|\xi\|_\infty \leq 1$, such that $(\xi, Du) = |Du|$ and

$$-\operatorname{div}\xi = u \quad \text{in } \mathcal{D}'(\mathbb{R}^N).$$

If the vector field $\xi \in L^\infty(\mathbb{R}^N; \mathbb{R}^N)$ is such that $\operatorname{div}\xi \in L^2(\mathbb{R}^N)$ and $u \in \mathrm{BV}(\mathbb{R}^N)$, the expression (ξ, Du) is a distribution defined by the formula

$$\langle(\xi, Dw), \varphi\rangle := -\int_{R^N} w\,\varphi \operatorname{div}\xi\,dx - \int_{R^N} w\,\xi\cdot\nabla\varphi\,dx \quad \text{for all } \varphi \in C_0^\infty(\mathbb{R}^N).$$

Then (ξ, Du) is a Radon measure in $\mathbb{R}^N$ which coincides with $\xi\cdot\nabla u$ when $u \in L^2(\mathbb{R}^N) \cap W^{1,1}(\mathbb{R}^N)$ [11].

The following result is taken from [13] and it explains how can we derive from solutions of (6) data z for which the solution of (5) is explicit.

Proposition 2.2. *Let $u_i \in \mathrm{BV}(\mathbb{R}^N)$ be such that $\inf(|u_i|, |u_j|) = 0$, $i, j \in \{1, \ldots, m\}$, $i \neq j$. Assume that u_i and $\sum_{i=1}^m u_i$ are solutions of the eigenvalue problem* (6),

$i \in \{1, \dots, m\}$. Let $b_i \in R$, $i = 1, \dots, m$, $z := \sum_{i=1}^m b_i u_i$, and $\lambda > 0$. Then the solution u of the variational problem (5) *is* $u := \sum_{i=1}^m \operatorname{sign}(b_i)(|b_i| - \lambda^{-1})^+ u_i$.

Assume that $m = 1$ and $\bar{u}$ is a solution of (6). If $0 < \lambda^{-1} \le b$, then $u := a\bar{u}$ with $a = b - \lambda^{-1}$ is a solution of (5) for the datum $z = b\bar{u}$. Indeed, u satisfies the Euler–Lagrange equation of (5) which characterizes its unique solution:

$$z = b\bar{u} = a\bar{u} + \lambda^{-1}\bar{u} = a\bar{u} - \lambda^{-1}\operatorname{div}\left(\frac{D\bar{u}}{|D\bar{u}|}\right) = u - \lambda^{-1}\operatorname{div}\left(\frac{Du}{|Du|}\right).$$

If $\lambda^{-1} > b$, then $u = 0$ is the solution of (5). Indeed, in this case $\|\lambda z\|_{\mathrm{BV}^*} \le 1$ (the dual norm in $\mathrm{BV}(\mathbb{R}^N)^*$) and there is a vector field $\xi \in L^\infty(\mathbb{R}^N; \mathbb{R}^N)$ with $\|\xi\|_\infty \le 1$, such that $-\operatorname{div}\xi = \lambda z$. Thus, $u = 0$ satisfies the Euler–Lagrange equation of (5). The proof when $b \le 0$ is similar and we skip the details. This solution exhibits a loss of contrast of size $\min(\lambda^{-1}, |b|)$ when the datum is $z = b\bar{u}$.

Our next theorem gives a family of solutions of (6) and is taken from [12] (see also [4]).

Theorem 2.3. *Let $C_1, \dots, C_m$ be bounded convex subsets of $\mathbb{R}^2$ which are disjoint. Let $b_i > 0$, $i = 1, \dots, m$, $k \in \{1, \dots, m\}$. Then $v := -\sum_{i=1}^k b_i \chi_{C_i} + \sum_{i=k+1}^m b_i \chi_{C_i}$ is a solution of* (6) *if and only if the following conditions holds.*

(i) *The sets C_i, $i = 1, \dots, m$, are of class $C^{1,1}$.*

(ii) *$b_i = \frac{P(C_i)}{|C_i|}$ for any $i \in \{1, \dots, m\}$.*

(iii) *The following inequalities hold:*

$$\operatorname*{ess\,sup}_{p \in \partial C_i} \kappa_{C_i}(p) \le \frac{P(C_i)}{|C_i|} \quad \text{for all } i = 1, \dots, m.$$

(iv) *If E_1 is a solution of the variational problem*

$$\min\left\{P(E) : \bigcup_{j=k+1}^m C_j \subseteq E \subseteq \mathbb{R}^2 \setminus \bigcup_{i=1}^k C_i\right\},$$

then we have

$$P(E_1) = \sum_{j=k+1}^m P(C_j).$$

If E_2 is a solution of the variational problem

$$\min\left\{P(E) : \bigcup_{i=1}^k C_i \subseteq E \subseteq \mathbb{R}^2 \setminus \bigcup_{j=k+1}^m C_j\right\},$$

then we have

$$P(E_2) = \sum_{i=1}^k P(C_i).$$

Moreover, if $k = m$, then we do not need to assume that the C_i are convex and we can replace condition (i) *by the following one:*

(i′) *The sets C_i, $i = 1, \dots, m$, are convex and of class $C^{1,1}$.*

This result was essentially proved in [12] (though we only stated the result in its second assertion). Its extension to $\mathbb{R}^N$ was proved in [4] (replacing the curvature of the boundaries by the sum of principal curvatures) under the assumption that the sets C_i are convex and of class $C^{1,1}$. Let us point out the following corollary for connected sets.

Corollary 2.4. *Let $C \subset \mathbb{R}^2$ be a bounded set of finite perimeter, and assume that C is connected. The function $v := \lambda \chi_C$ is a solution of* (6) *if and only if the following three conditions hold.*

(i) $\lambda = \lambda_C := \frac{P(C)}{|C|}$.

(ii) *C is convex and ∂C is of class $C^{1,1}$.*

(iii) *The following inequality holds:*

$$\operatorname*{ess\,sup}_{p \in \partial C} \kappa_{\partial C}(p) \leq \frac{P(C)}{|C|}.$$

A convex set $C \subseteq \mathbb{R}^2$ such that $u := \lambda_C \chi_C$ is a solution of (6) is called calibrable. The above result gives a characterization of calibrable sets in $\mathbb{R}^2$ and was proved in [24], [12]. For convex sets in $\mathbb{R}^N$ of class $C^{1,1}$ the above result is true if we replace the curvature of the boundary by the sum of the principal curvatures [4].

Example 1. Let $C \subset \mathbb{R}^2$ be the set of Figure 1. It is easy to check that C satisfies the

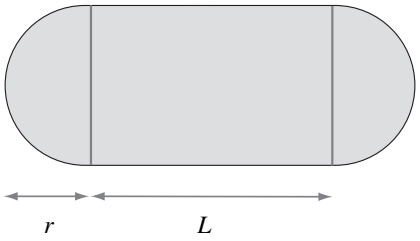

Figure 1. A bean-shaped set is calibrable.

assumptions of Corollary 2.4, since C is a convex set with $C^{1,1}$ boundary and there holds

$$\operatorname*{ess\,sup}_{p \in \partial C} \kappa_{\partial C}(p) = \frac{1}{r} < \frac{2\pi r + 2L}{\pi r^2 + 2rL} = \frac{P(C)}{|C|}. \tag{7}$$

Moreover, since the inequality in (7) is always strict, any convex set C' of class $C^{1,1}$ close enough to C in the $C^{1,1}$-norm is also calibrable.

Example 2. Let $\Omega \subset \mathbb{R}^2$ be the union of two disjoint balls B_1 and B_2 of radius r, whose centers are at distance L (see Figure 2). Then $k = 0$ and $m = 2$ in Theorem 2.3 and condition (iv) in it reads as

$$L \geq \pi r.$$

Under this condition the set Ω is calibrable. The condition $L \geq \pi r$ is nothing else than $P(\mathrm{co}(B_1 \cup B_2)) \geq P(B_1) + P(B_2)$ (co denotes the convex envelope) and in this case the solution of the denoising problem with $z = \chi_{B_1 \cup B_2}$ coincides with the

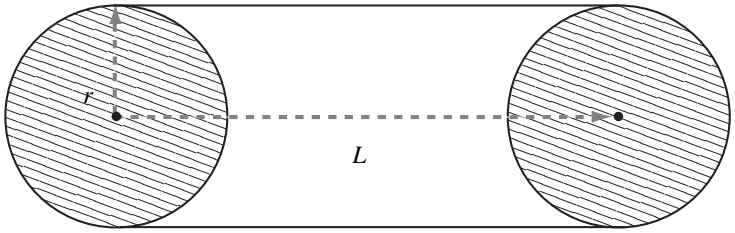

Figure 2. Two balls as initial datum for the denoising problem.

addition of the solutions obtained with χ_{B_1} and χ_{B_2}. In case that $P(\mathrm{co}(B_1 \cup B_2)) < P(B_1) + P(B_2)$ there is interaction of the two sets and the solution is not the addition of solutions corresponding to the data χ_{B_1} and χ_{B_2}.

These solutions exhibit two features of (5): that discontinuities may be preserved and the loss of contrast.

We could expand the above family of solutions by classifying all possible solutions of (6). Along this line, we extended the above results in two directions: in [13] we looked for solutions of (6) which are built up as sums of linear combinations of characteristic functions of convex sets of class $C^{1,1}$ (not disjoint, in general), and we considered in [3], [4] the case of general convex sets.

Let us illustrate the results in [13] with a simple case.

Proposition 2.5. *Let K_0, K_1 be two bounded open convex sets of $\mathbb{R}^2$ with boundary of class $C^{1,1}$ such that $\overline{K_1} \subseteq K_0$. Let $F := K_0 \setminus \overline{K_1}$. Let*

$$J := \frac{P(K_0) - P(K_1)}{|F|} > 0.$$

If

$$\operatorname*{ess\,sup}_{\partial K_0} \kappa_{\partial K_0} \leq J, \quad \operatorname*{ess\,inf}_{x \in \partial K_1} \kappa_{\partial K_1}(x) \geq J, \quad \operatorname*{ess\,sup}_{x \in \partial K_1} \kappa_{\partial K_1}(x) \leq \lambda_{K_1}$$

then $v = \lambda_{K_1} \chi_{K_1} + J \chi_{K_0 \setminus K_1}$ is a solution of (6).

The works [3], [4] describe the denoising of the characteristic function of any convex set of $\mathbb{R}^2$ and $\mathbb{R}^N$, respectively, and the results in them illustrate the regularization of corners. Even if the more general case of linear combinations of convex sets in $\mathbb{R}^2$ and $\mathbb{R}^N$ is considered, we illustrate the results in [3], [4] with a simple case.

Theorem 2.6. *Assume that C is a bounded convex set in $\mathbb{R}^2$. Then there is a calibrable set $C_R \subseteq C$ such that $\partial C \setminus \partial C_R$ is formed by arcs of circle of radius R such that $\frac{1}{R} = \frac{P(C_R)}{|C_R|}$ and for each $x \in C \setminus C_R$ it passes a unique arc of circle of radius $r(x)$ and those circles fiber $C \setminus C_R$. Let $r(x) = R$ for $x \in C_R$. Then $u(x) = \left(1 - \frac{\lambda^{-1}}{r(x)}\right)^+ \chi_C$ is the solution of* (6) *for the data $z = \chi_C$.*

3. Image decomposition models

In his work [26], Y. Meyer interpreted the denoising model as a $u + v$ decomposition. Assume that Ω is a bounded connected domain in $\mathbb{R}^2$ with Lipschitz boundary. If $z \in L^2(\Omega)$ and u is the solution of (5), then its Euler–Lagrange equation can be written as

$$u + v = z \quad \text{where } v = -\frac{1}{\lambda}\operatorname{div}\left(\frac{Du}{|Du|}\right).$$

This type of decompositions is called a $u + v$ decomposition and u is supposed to be a geometric sketch of the image [26]. As we have shown in the previous section, model (5) does not attain its objective of separating an image into its $u + v$ decomposition. This conclusion was also derived in [26] through complementary arguments. For instance, if $z = \chi_\omega$ where ω is a bounded domain with a C^∞ boundary, then z is not preserved by the Rudin–Osher–Fatemi (ROF) model (contrary to what it should be expected). The v component contains the noise but also part of the image structure and, in particular, part of the texture (depending on the value of λ). On the other hand if $z(x) = \chi_A(x) + p(mx)\chi_B(x)$ where A and B are two bounded domains with smooth boundary, $m \geq 1$, and $p(x)$, $x = (x_1, x_2)$, is a smooth 2π-periodic function of the two variables x_1, x_2, then the ROF model does not give $u(x) = \chi_A(x)$, $v(x) = p(mx)\chi_B(x)$ [26] (this will be explained after Theorem 3.1). Then to improve the ROF model Meyer proposed a different decomposition [26], which is based in the following variational model

$$\inf_{u \in \mathrm{BV}(\Omega), v \in G(\Omega), z = u+v} \int_\Omega |Du|\,dx + \lambda \|v\|_G,$$

where $\lambda > 0$ and $G(\Omega)$ denotes the Banach space of distributions f in Ω that may be written

$$f = \operatorname{div} \xi$$

where $\xi \in L^\infty(\Omega; \mathbb{R}^2)$. The norm in G is defined by

$$\|f\|_G := \inf\{\|\xi\|_\infty : \xi \in L^\infty(\Omega; \mathbb{R}^2),\ f = \operatorname{div}\xi\}$$

where $\|\xi\|_\infty := \operatorname{ess\,sup}_{x\in\Omega} |\xi(x)|$. $G(\Omega)$ is exactly $W^{-1,\infty}(\Omega)$, the dual space of $W_0^{1,1}(\Omega)$. The justification for the introduction of the space G comes from the next result [26].

Theorem 3.1. *Let f_n be a sequence of functions in $L^2(\Omega)$ with the following properties*

(i) *There exists a compact set $K \subset \Omega$ such that the support of f_n is contained in K for each n,*

(ii) *There exists $q > 2$ and $C > 0$ such that $\|f_n\|_q \leq C$,*

(iii) *The sequence f_n converges to 0 in a distributional sense.*

Then $\|f_n\|_G$ converges to 0 as $n \to \infty$.

In other words, oscillating textures have a small norm in $G(\Omega)$. Now, if $z(x) = \chi_A(x) + p(mx)\chi_B(x)$ is as in the first paragraph of this section, then v cannot be $p(mx)\chi_B(x)$ for large m [26]. Otherwise we would have $p(mx)\chi_B(x) = -\frac{1}{\lambda}\operatorname{div}\left(\frac{Du}{|Du|}\right)$ and therefore $\|p(mx)\chi_B(x)\|_G = \frac{1}{\lambda}$. But we know from Theorem 3.1 that the G-norm of $p(mx)\chi_B(x)$ is small for large values m (indeed the G-norm of $p(mx)\chi_B(x)$ is an $O(m^{-1})$ [26]).

Theorem 3.1 and other results [26], [25] were the starting point of extensive numerical work on $u + v$ decompositions [32], [28], [10], [9] to explore and compare the relative ability of the G based model versus the ROF model. Meyer's model was first implemented by Vese–Osher in [32]. A different approach was proposed in [10], [9] where the decomposition is computed by minimizing a convex functional which depends on the two variable u and v, alternatively in each variable. Each minimization is based on the projection algorithm introduced in [16]. The problem to solve is:

$$\inf_{(u,v)\in \mathrm{BV}(\Omega)\times\mu B_G} \int_\Omega |Du| + \frac{\lambda}{2}\int_\Omega |z-u-v|^2\,dx, \tag{8}$$

where $B_G := \{v \in G : \|v\|_G \leq 1\}$. We refer to [10] for its precise connection with Meyer's model. Let us mention that other dual Sobolev norms, indeed H^{-1}, have been explored in [28].

Figure 3 displays the comparison between ROF and model (8) for a simple figure. These images are courtesy of J. F. Aujol and A. Chambolle and have been obtained with the numerical methods developed by the authors in [9], [10]. Figures 3.a and 3.b display the original reference image and the noisy image with an additive Gaussian white noise with $\sigma = 35$. Figures 3.c and 3.d display the u and v components obtained with the ROF model with λ chosen so that $\|v\| = \sigma$. For better visualization, the v component will be always displayed as $v + 128$. Figures 3.e and 3.f display the u and v components obtained with model (8) with $\lambda = 10$ and $\mu = 55$ (for more details on the choice of parameters, see [9], [10]). In this case, for well chosen values of the parameter, the results are quite comparable. But let us point out that model (8) is able

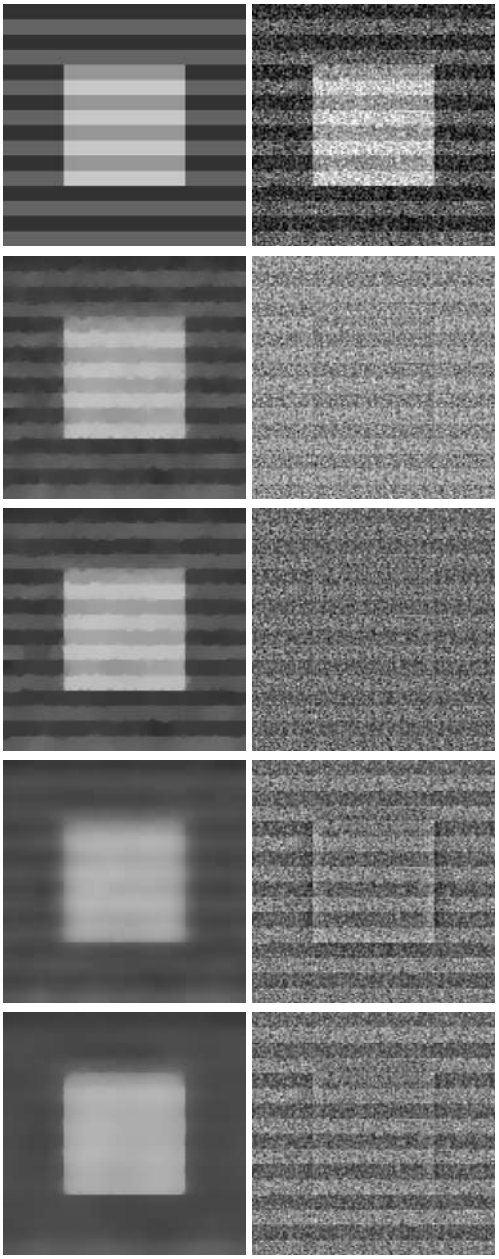

Figure 3. *Comparison of ROF and model* (8). From left to right and top to bottom: a) Original reference image. b) Noisy image with $\sigma = 35$. c) and d) Result of the ROF model: u and v component (with λ chosen so that $\|v\| = \sigma$). For a better visualization, the v component will be displayed as $v + 128$. e) and f) Result of model (8): u and v component ($\lambda = 10$, $\mu = 55$). g) and h) Result of the ROF model: u and v component (in this case $\sigma = 40.8$). i) and j) Result of model (8): u and v component (in this case $\sigma = 40.8$ and $\mu = 200$). These images are courtesy of J. F. Aujol and A. Chambolle. See the text for more details.

to separate the horizontal bands from the square for large values of the parameter μ while this does not seem to be possible with the ROF model. This is displayed in the next figures. Figures 3.g and 3.h display the u and v components obtained with the ROF model (the noise corresponds to a value of $\sigma = 40.8$) with λ chosen so that $\|v\| = \sigma$. Figures 3.i and 3.j display the u and v components obtained with model (8) with $\mu = 200$. In any case, the choice of the parameters is open to further analysis and it the separation of the image in two components is related to the different scales present in the image.

4. Image restoration

To approach the problem of image restoration from a numerical point of view we shall assume that the image formation model incorporates the sampling process in a regular grid

$$z(i, j) = h * u(i, j) + n(i, j), \quad (i, j) \in \{1, \dots, N\}^2 \tag{9}$$

where $u\colon \mathbb{R}^2 \to \mathbb{R}$ denotes the ideal undistorted image, $h\colon \mathbb{R}^2 \to \mathbb{R}$ is a blurring kernel, z is the observed sampled image which is represented as a function $z\colon \{1, \dots, N\}^2 \to \mathbb{R}$, and $n(i, j)$ is, as usual, a white Gaussian noise with zero mean and standard deviation σ.

Let us denote by Ω_N the interval $(0, N]^2$. As we said in the introduction, in order to simplify this problem, we assume that the functions h and u are periodic of period N in each direction. That amounts to neglecting some boundary effects. Therefore, we assume that h, u are functions defined in Ω_N. To fix ideas, we assume that $h, u \in L^2(\Omega_N)$, so that $h * u$ is a continuous function in Ω_N (which may be extended to a continuous periodic function in $\mathbb{R}^2$) and the samples $h * u(i, j)$, $(i, j) \in \{1, \dots, N\}^2$, have sense.

Our next purpose is to introduce a restoration model with local constraints and to explain the numerical approach to solve it. For that, let us introduce some notation. We denote by X the Euclidean space $\mathbb{R}^{N\times N}$. Then the image $u \in X$ is the vector $u = (u(i, j))_{i,j=1}^N$, and the vector field ξ is the map $\xi\colon \{1, \dots, N\} \times \{1, \dots, N\} \to \mathbb{R}^2$. If $u \in X$, the discrete gradient is a vector in $Y = X \times X$ given by

$$\nabla^{+,+}u := (\nabla_x^+ u, \nabla_y^+ u),$$

where

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

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

for $i, j \in \{1, \dots, N\}$. We denote $\nabla^{+,+}u = (\nabla_x^+ u, \nabla_y^+ u)$. Other choices of the gradient are possible, this one will be convenient for the developments below.

Let us define the discrete functional

$$J_d^\beta(u) = \sum_{1 \le i,j \le N} \sqrt{\beta^2 + |\nabla^{+,+}u(i,j)|^2}, \quad \beta \ge 0.$$

For any function $w \in L^2(\Omega_N)$, its Fourier coefficients are

$$\hat{w}_{\frac{l}{N},\frac{j}{N}} = \int_{\Omega_N} w(x,y) e^{-2\pi i \frac{(lx+jy)}{N}} \quad \text{for } (l,j) \in \mathbb{Z}^2.$$

Our plan is to compute a band limited approximation to the solution of the restoration problem for (9). For that we define

$$\mathcal{B} := \left\{ u \in L^2(\Omega_N) : \hat{u} \text{ is supported in } \{-\tfrac{1}{2} + \tfrac{1}{N}, \dots, \tfrac{1}{2}\} \right\}.$$

We notice that $\mathcal{B}$ is a finite dimensional vector space of dimension N^2 which can be identified with X. Both $J(u) = \int_{\Omega_N} |Du|$ and $J_d^0(u)$ are norms on the quotient space $\mathcal{B}/\mathbb{R}$, hence they are equivalent. With a slight abuse of notation we shall indistinctly write $u \in \mathcal{B}$ or $u \in X$.

We shall assume that the convolution kernel $h \in L^2(\Omega_N)$ is such that $\hat{u}$ is supported in $\left\{ -\frac{1}{2} + \frac{1}{N}, \dots, \frac{1}{2} \right\}$ and $\hat{h}(0,0) = 1$.

In the discrete framework, the ROF model for restoration is

$$\text{Minimize}_{u \in X} \quad J_d^\beta(u) \tag{10}$$

$$\text{subject to} \quad \sum_{i,j=1}^{N} |h * u(i,j) - z(i,j)|^2 \le \sigma^2 N^2. \tag{11}$$

Notice again that the image acquisition model (1) is only incorporated through a global constraint. In practice, the above problem is solved via the following unconstrained formulation

$$\min_{u \in X} \max_{\lambda \ge 0} J_d^\beta(u) + \frac{\lambda}{2}\Big[\frac{1}{N^2} \sum_{i,j=1}^{N} |h * u(i,j) - z(i,j)|^2 - \sigma^2 \Big] \tag{12}$$

where $\lambda \ge 0$ is a Lagrange multiplier. The appropriate value of λ can be computed using Uzawa's algorithm [15], [2] so that the constraint (11) is satisfied. Recall that if we interpret λ^{-1} as a penalization parameter which controls the importance of the regularization term, and we set this parameter to be small, then homogeneous zones are well denoised while highly textured regions will loose a great part of its structure. On the contrary, if λ^{-1} is set to be small, texture will be kept but noise will remain in homogeneous regions. On the other hand, as the authors of [15], [2] observed, if we use the constrained formulation (10)-(11) or, equivalently (12), then

the Lagrange multiplier does not produce satisfactory results since we do not keep textures and denoise flat regions simultaneously, and they proposed to incorporate the image acquisition model as a set of local constraints.

Following [2], we propose to replace the constraint (11) by

$$G*(h*u-z)(i,j) \le \sigma^2, \quad \text{for all } (i,j) \in \{1,\dots,N\}, \tag{13}$$

where G is a discrete convolution kernel such that $G(i,j) > 0$ for all $(i,j) \in \{1,\dots,N\}$. The effective support of G must permit the statistical estimation of the variance of the noise with (13) (see [2]). Then we shall minimize the functional $J_d^\beta(u)$ on X submitted to the family of constraints (13) (plus eventually the constraint $\sum_{i,j=1}^N h*u(i,j) = \sum_{i,j=1}^N z(i,j)$). Thus, we propose to solve the optimization problem:

$$\begin{aligned} &\min_{u\in\mathcal{B}} J_d^\beta(u) \\ &\text{subject to} \quad G*(h*u-z)^2(i,j) \le \sigma^2 \quad \text{for all } (i,j). \end{aligned} \tag{14}$$

This problem is well-posed, i.e., there exists a solution and is unique if $\beta > 0$ and $\inf_{c\in\mathbb{R}} G*(z-c)^2 > \sigma^2$. In case that $\beta = 0$ and $\inf_{c\in\mathbb{R}} G*(z-c)^2 > \sigma^2$, then $h*u$ is unique. Moreover, it can be solved with a gradient descent approach and Uzawa's method [2].

To guarantee that the assumptions of Uzawa's method hold we shall use a gradient descent strategy. For that, let $v \in X$ and $\gamma > 0$. At each step we have to solve a problem like

$$\begin{aligned} &\min_{u\in\mathcal{B}} \gamma|u-v|_X^2 + J_d^\beta(u) \\ &\text{subject to} \quad G*(h*u-z)^2(i,j) \le \sigma^2 \quad \text{for all } (i,j). \end{aligned} \tag{15}$$

We solve (15) using the unconstrained formulation

$$\min_{u\in X} \max_{\lambda\ge 0} \mathcal{L}^\gamma(u,\{\lambda\};v),$$

where $\lambda = (\lambda(i,j))_{i,j=1}^N$ and

$$\mathcal{L}^\gamma(u,\{\lambda\};v) = \gamma|u-v|_X^2 + J_d^\beta(u) + \sum_{i,j=1}^N \lambda(i,j)(G*(h*u-z)^2(i,j)-\sigma^2).$$

Algorithm: TV based restoration algorithm with local constraints

1. Set $u^0 = 0$ or, better, $u^0 = z$. Set $n = 0$.
2. Use Uzawa's algorithm to solve the problem

$$\min_{u\in X} \max_{\lambda\ge 0} \mathcal{L}^\gamma(u,\{\lambda\};u^n), \tag{16}$$

that is:

(a) Choose any set of values $\lambda^0(i,\, j) \geq 0$, $(i,\, j) \in \{1, \ldots, N\}^2$, and $u_0^n = u^n$. Iterate from $p = 0$ until convergence of λ^p the following steps:

(b) With the values of λ^p solve the problem

$$\min_u \mathcal{L}^\gamma(u, \{\lambda^p\}; u^n)$$

starting with the initial condition u_p^n. Let u_{p+1}^n be the solution obtained.

(c) Update λ in the following way:

$$\lambda^{p+1}(i,\, j) = \max(\lambda^p(i,\, j) + \rho(G * (h * u_p^n - z)^2(i,\, j) - \sigma^2), 0)$$

for all $(i,\, j)$.

Let u^{n+1} be the solution of (16). Stop when convergence of u^n.

We notice that, since $\gamma > 0$, Uzawa's algorithm converges if $z \in h * \mathcal{B}$. Moreover, if u^0 satisfies the constraints, then u^n tends to a solution u of (14) as $n \to \infty$ [2].

Finally, to solve problem (16) in Step 2.(b) of the algorithm we use either the extension of Chambolle's algorithm [16] to the restoration case given in [1] if we use $\beta = 0$, or the Bermúdez–Moreno algorithm [14] adapted to solve (16) when $\beta > 0$ as given in [2]. Being differentiable at when $\nabla^{+,+}u = 0$, this second possibility produces slightly smoother solutions in smooth non textured areas. We shall not enter on the comparison of both possibilities here and we shall use $\beta = 0$. For more details, we refer to [1], [2].

Let us mention the work [23] where the authors introduce a spatially varying fidelity term which controls the amount of denoising in any region of the image in order to preserve textures and small details. The philosophy is the same as ours but the value of $\lambda(i,\, j)$ is chosen in a different way.

5. Some restoration experiments

To simulate our data we use the modulation transfer function corresponding to SPOT 5 HRG satellite with Hipermode sampling (see [29] for more details):

$$\hat{h}(\eta_1, \eta_2) = e^{-4\pi\beta_1|\eta_1|} e^{-4\pi\alpha\sqrt{\eta_1^2+\eta_2^2}} \operatorname{sinc}(2\eta_1) \operatorname{sinc}(2\eta_2) \operatorname{sinc}(\eta_1), \qquad (17)$$

where $\eta_1, \eta_2 \in [-1/2, 1/2]$, $\operatorname{sinc}(\eta_1) = \sin(\pi\eta_1)/(\pi\eta_2)$, $\alpha = 0.58$, and $\beta_1 = 0.14$. Then we filter the reference image given in Figure 4.a with the filter (17) and we add some Gaussian white noise of zero mean and standard deviation σ (in our case $\sigma = 1$, which is a realistic assumption for the case of satellite images [29]) to obtain the image displayed in Figure 4.b.

Figure 5.a displays the restoration of the image in Figure 4.b obtained using the algorithm of last section with $\beta = 0$. We have used a Gaussian function G of standard deviation $\sigma = 6$. The mean value of the constraint is $\text{mean}((G*(Ku-z))^2) = 1.0933$ and RMSE $= 7.9862$. Figure 5.b displays the function $\lambda(i,\, j)$ obtained.

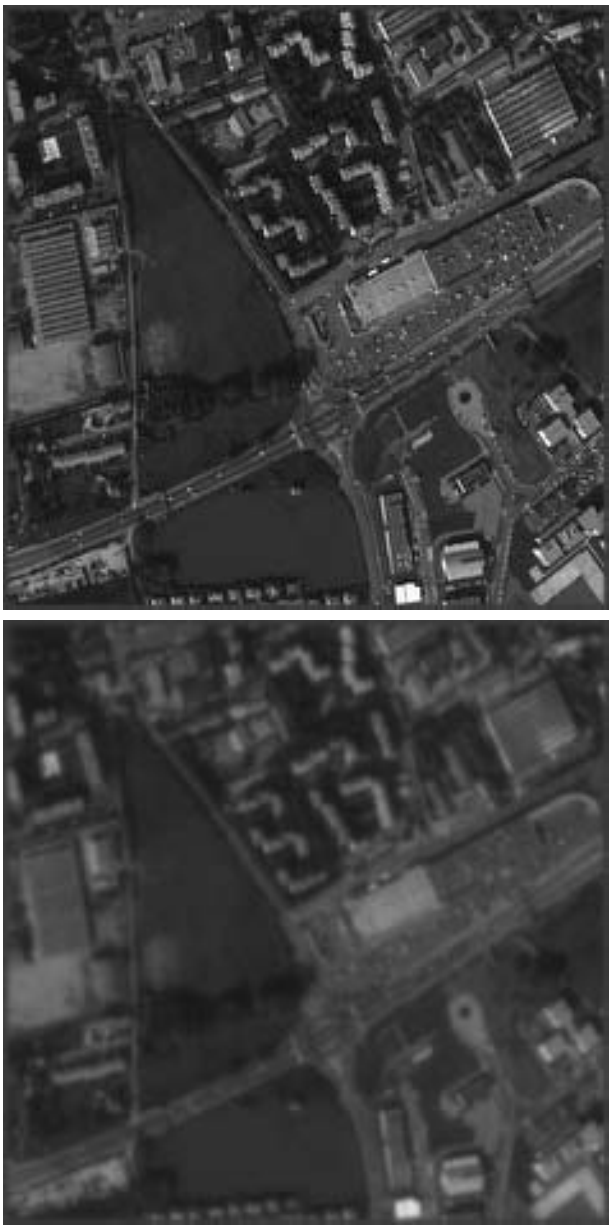

Figure 4. *Reference image and a filtered and noised image.* a) Top: reference image. b) Bottom: the data. This image has been generated applying the MTF given in (17) to the top image and adding a Gaussian white noise of zero mean and standard deviation $\sigma = 1$.

Figure 6 displays some details of the results that are obtained using a single global constraint (11) and show its main drawbacks. Figure 6.a corresponds to the result obtained with the Lagrange multiplier $\lambda = 10$ (thus, the constraint (11) is satisfied). The result is not satisfactory because it is difficult to denoise smooth regions and keep the textures at the same time. Figure 6.b shows that most textures are lost when using a small value of λ ($\lambda = 2$) and Figure 6.c shows that some noise is present if we use a larger value of λ ($\lambda = 1000$). This result is to be compared with the same detail of Figure 5.a which is displayed in Figure 6.d.

The modulation transfer function for satellite images. We describe here a simple model for the Modulation Transfer Function of a general satellite. More details can be found in [29] where specific examples of MTF for different acquisition systems are shown. The MTF used in our experiments (17) corresponds to a particular case of the general model described below [29].

Recall that the MTF, that we denote by $\hat{h}$, is the Fourier transform of the impulse response of the system. Let $(\eta_1, \eta_2) \in [-1/2, 1/2]$ denote the coordinates in the frequency domain. There are different parts in the acquisition system that contribute to the global transfer function:

Sensors. Every sensor has a sensitive region where all the photons that arrive are integrated. This region can be approximated by a unit square $[-c/2, c/2]^2$ where c is the distance between consecutive sensors. Its impulse response is then the convolution of two pulses, one in each spatial direction. The corresponding transfer function also includes the effect of the conductivity (diffusion of information) between neighbouring sensors, which is modeled by an exponential decay factor, thus:

$$\hat{h}_S(\eta_1, \eta_2) = \operatorname{sinc}(\eta_1 c)\operatorname{sinc}(\eta_2 c) e^{-2\pi\beta_1 c|\eta_1|} e^{-2\pi\beta_2 c|\eta_2|},$$

where $\operatorname{sinc}(\eta_1) = \sin(\pi\eta_1)/(\pi\eta_1)$ and $\beta_1, \beta_2 > 0$.

Optical system. It is considered as an isotropic low-pass filter

$$\hat{h}_O(\eta_1, \eta_2) = e^{-2\pi\alpha c\sqrt{\eta_1^2+\eta_2^2}}, \quad \alpha > 0.$$

Motion. Each sensor counts the number of photons that arrive to its sensitive region during a certain time of acquisition. During the sampling time the system moves a distance τ and so does the sensor; this produces a motion blur effect in the motion direction (d_1, d_2):

$$\hat{h}_M(\eta_1, \eta_2) = \operatorname{sinc}(\langle(\eta_1, \eta_2), (d_1, d_2)\rangle\tau).$$

Finally, the global MTF is the product of each of these intermediate transfer functions modeling the different aspects of the satellite:

$$\hat{h}(\eta_1, \eta_2) = \hat{h}_S\hat{h}_O\hat{h}_M.$$

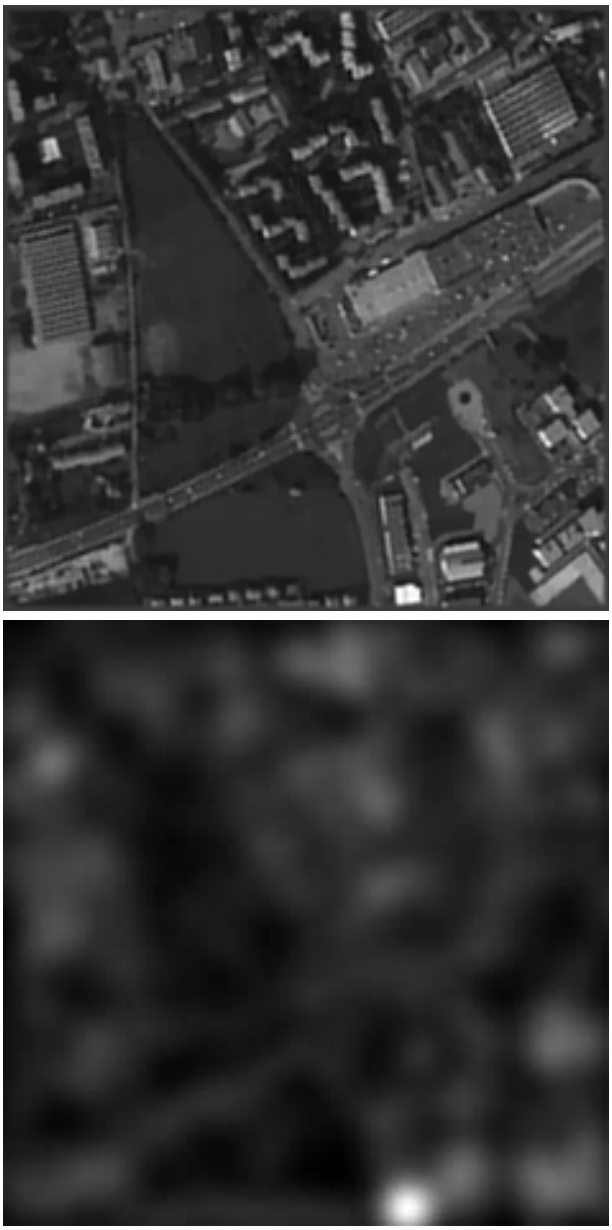

Figure 5. *Restored image with local Lagrange multipliers.* a) Top: the restored image corresponding to the data given in Figure 4.b. The restoration has been obtained using the algorithm of last section We have used a Gaussian function G of standard deviation $\sigma = 6$. b) Bottom: the function $\lambda(i, j)$ obtained.

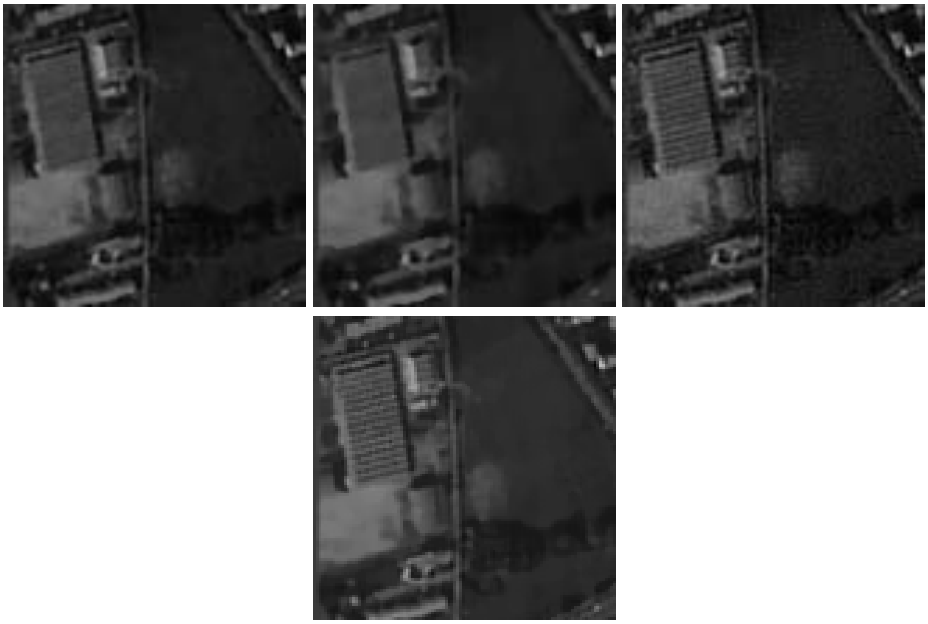

Figure 6. *A detail of the restored images with global and local constraints.* Top: a), b) and c) display a detail of the results that are obtained using a single global constraint (11) and show its main drawbacks. Figure a) corresponds to the result obtained with the value of λ such that the constraint (11) is satisfied, in our case $\lambda = 10$. Figure b) shows that most textures are lost when using a small value of λ ($\lambda = 2$)and Figure c) shows that some noise is present if we use a larger value of λ ($\lambda = 1000$). Bottom: d) displays the same detail of Figure 5.a which has been obtained using restoration with local constraints.

Acknowledgements. The author is indebted to his coauthors: A. Almansa, F. Alter, F. Andreu, C. Ballester, G. Bellettini, M. Bertalmío, A. Chambolle, J. Faro, G. Haro, J. M. Mazón, M. Novaga, B. Rougé, and A. Solé. I would like to thank also J. F. Aujol and A. Chambolle for providing me with the images in Figure 3. Special thanks and a warm dedication to Andrés Solé who passed away after completing his PhD and started with me his work on restoration.

References

[1] Almansa, A., Caselles, V., Haro, G., and Rougé, B., Restoration and zoom of irregularly sampled, blurred and noisy images by accurate total variation minimization with local constraints. *Multiscale Model. Simul.*, to appear.

[2] Almansa, A., C. Ballester, C., Caselles, V., Faro, L. J., and Haro, G., A TV based restoration model with local constraints. Preprint, 2006.

[3] Alter, F., Caselles, V., and Chambolle, A., Evolution of Characteristic Functions of Convex Sets in the Plane by the Minimizing Total Variation Flow. *Interfaces Free Bound.* **7** (2005), 29–53.

[4] Alter, F., Caselles, V., and Chambolle, A., A characterization of Convex Calibrable Sets in $\mathbb{R}^N$. *Math. Ann.* **332** (2005), 329–366.

[5] Ambrosio, L., Fusco, N., and Pallara, D., *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Math. Monogr., The Clarendon Press, Oxford University Press, New York 2000

[6] Andreu, F., Ballester, C., Caselles, V., and Mazón, J. M., Minimizing total variation flow. *Differential Integral Equations* **14** (2001), 321–360.

[7] Andreu-Vaillo, F., Caselles, V., and Mazón, J. M., *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals.* Progr. Math. 223, Birkhäuser, Basel 2004

[8] Andrews, H. C., and Hunt, B. R., *Digital Image Restoration.* Prentice Hall, Englewood Cliffs, NJ, 1977.

[9] Aujol, J. F., and Chambolle, A., Dual Norms and Image Decomposition Models. *Internat. J. Computer Vision*, **63** (2005), 85–104.

[10] Aujol, J. F., Aubert, G., Blanc-Feraud, L., and Chambolle, A., Image Decomposition into a bounded variation component and an oscillating component. *J. Math. Imaging Vision* **22** (2005), 71–88.

[11] Anzellotti, G., Pairings between measures and bounded functions and compensated compactness. *Ann. Mat. Pura Appl.* **135** (1983), 293–318.

[12] Bellettini, G., Caselles, V., and Novaga, M., The Total Variation Flow in R^N. *J. Differential Equations* **184** (2002), 475–525.

[13] Bellettini, G., Caselles, V., and Novaga, M., Explicit solutions of the eigenvalue problem $-\mathrm{div}\left(\frac{Du}{|Du|}\right) = u$. *SIAM J. Math. Anal.* **36** (2005), 1095–1129.

[14] Bermúdez, A., and Moreno, C., Duality methods for solving variational inequalities. *Comput. Math. Appl.* **7** (1981), 43–58.

[15] Bertalmío, M., Caselles, V., Rougé, B., and Solé, A., TV based image restoration with local constraints. *J. Sci. Comput.* **19** (2003), 95–122.

[16] Chambolle, A., An algorithm for total variation minimization and applications. *J. Math. Imaging Vision* **20** (2004), 89–97.

[17] Chambolle, A., and Lions, P. L., Image recovery via total variation minimization and related problems. *Numer. Math.* **76** (1997), 167–188.

[18] Chan, T. F., Golub, G. H., and Mulet, P., A Nonlinear Primal-Dual Method for Total Variation Based Image Restoration. *SIAM J. Sci. Comput.* **20** (1999), 1964–1977.

[19] Demoment, G., Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Trans. Acoust. Speech Signal Process.* **37** (1989), 2024–2036.

[20] Donoho, D., Denoising via soft-threholding. *IEEE Trans. Inform. Theory* **41** (1995), 613–627.

[21] Durand, S., Malgouyres, F., and Rougé, B., Image Deblurring, Spectrum Interpolation and Application to Satellite Imaging. *ESAIM Control Optim. Calc. Var.* **5** (2000), 445–475.

[22] Geman, D., and Reynolds, G., Constrained Image Restoration and Recovery of Discontinuities. *IEEE Trans. Pattern Anal. Machine Intell.* **14** (1992), 367–383.

[23] Gilboa, G., Sochen, N., and Zeevi, Y., PDE-based denoising of complex scenes using a spatially-varying fidelity term. In *Proc. International Conference on Image Processing 2003, Barcelona, Spain*, Vol. 1, 2003, 865–868.

[24] Giusti, E., On the equation of surfaces of prescribed mean curvature. Existence and uniqueness without boundary conditions. *Invent. Math.* **46** (1978), 111–137.

[25] Haddad, A., and Meyer, Y., Variational methods in image processing. *CAM Reports* 04-52, 2004.

[26] Meyer, Y., *Oscillating patterns in image processing and in some nonlinear evolution equations*. The Fifteenth Dean Jacqueline B. Lewis memorial lectures, University Lecture Series 22, Amer. Math. Soc., Providence, RI, 2001.

[27] Nikolova, M., Local strong homogeneity of a regularized estimator. *SIAM J. Appl. Math* **61** (2000), 633–658.

[28] Osher, S. J., Sole, A., and Vese, L. A., Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Model. Simul.* **1** (2003), 349–370.

[29] Rougé, B., Théorie de l'echantillonage et satellites d'observation de la terre. In *Analyse de Fourier et traitement d'images*, Journées X-UPS 1998.

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

[31] Tikhonov, A. N., and Arsenin, V. Y., *Solutions of Ill-Posed Problems*. Scripta Series in Mathematics, John Wiley & Sons, New York 1977.

[32] Vese, L. A., and Osher, S. J., Modeling textures with total variation mkinimization and oscillatikng patterns in image processing. *J. Sci. Comput.* **19** (2003), 553–572.

Departament Tecnologia, Universitat Pompeu Fabra, Passeig de Circumvalació, 8,
08003 Barcelona, Spain

E-mail: vicent.caselles@upf.edu