Image Denoising with a Constrained Discrete Total Variation Scale Space

Igor Ciril¹ and Jérôme Darbon²

¹ LMCS, IPSA

² CMLA, ENS Cachan, CNRS, PRES UniverSud

Abstract. This paper describes an approach for performing image restoration using a coupled differential system that both simplifies the image while preserving its contrast. The first process corresponds to a differential inclusion involving discrete Total Variations that simplifies more and more the observed image as time evolves. The second one extracts some pertinent geometric information contained in the series of simplified images and recovers the contrast using Bregman distances. Convergence and exact computational properties of the method rely on the discrete and combinatorial properties of discrete Total Variations.

Keywords: Discrete Total Variation, Bregman Distances, Differential Inclusions, Network Flows.

1 Introduction

Minimization of the Total Variation (TV) with a quadratic data fidelity term is a popular tool for performing image restoration since the seminal work of [3,23]. Although the solution has sharp boundaries it is also known that the minimizer may present a loss of contrast [17,24]. In this paper we propose an approach to address this issue within the framework of a coupled scale-space process.

One popular approach to avoid the loss of contrast consists in considering robust edge-preserving priors [1,5,8,12,18,19,25] instead of TV. Such regularization terms aim at not penalizing too much large gradient that are assumed to correspond to a contour in the reconstructed image. Among these priors, the Potts model and truncated-quadratic prior are the most well-known. However, such a prior yields a non-convex optimization problems where a global minimizer cannot be generally computed in practice. Thus a local minimum or a critical point is obtained using an approximation algorithm [8].

Another approach relies on a Bregman distance procedure that is originally proposed by Osher et al. in [20]. This scheme is an iterative method that consists of minimizing a sequence of convex minimization problems where each of them refines at each step a degraded image. More precisely, the process starts from a constant image and converges toward the observed noisy image. It relies on iterating the following two steps: the normals of the levels are firstly filtered using TV while a surface is approximately fitted to these estimated normal in a second step. These two operations are iterated until the reconstructed image satisfies a

prescribe criterion. The contrast of the reconstructed image is much better than the one obtained from a pure TV approach since it emphasized the geometric information contained in the normals of the level lines as opposed to the actual gray values. This procedure has also received the name of inverse scale space [4] since objects of finer scale are incrementally added to the reconstructed image. Note that this approach can also be rewritten in terms of Bregman distance [20]. In [10], some theoretical links between the iterative Bregman distance scheme and the use of robust edge preserving priors are highlighted.

In this paper we consider an approach that relies on coupling the TV-flow (which corresponds to the solution of a differential inclusion) that incrementally simplifies the original noisy image with a procedure that intends to recover the contrast. More precisely, the flow is used to extract some information about the geometry of the image: in this paper we consider the relative order between any two adjacent pixels (note that this information essentially corresponds to considering the normal of the level lines as in the Bregman process). This information is used in a second step that looks for an image that is the closest (relatively to l^2) to the observed data while it has the same relative order as the one obtained from the flow image. Interestingly enough, the imposition of the relative order is achieved through a Bregman distance. Note that our approach is a forward scale-space approach (as opposed to inverse as in [20]).

The remainder of this paper is as follows: some useful notations and definitions are given in section 2 while some theoretical results about differential inclusions are presented in section 3. Our approach is described in section 4. Finally we draw some conclusions in section 5.

2 Some Notations and Definitions

A discrete Markovian framework is considered [25]. An image is defined on a lattice referred to as $\mathcal{V}$ with cardinality $|\mathcal{V}| = N$. We see any image as a vector living in $\mathbb{R}^N$. The value of an image $\mathbf{u}$ at the site $i \in \mathcal{V}$ is referred to as u_i . We endow the lattice with a neighborhood system and we consider pairwise interactions between sites. We denote by (i, j) the interaction between two sites i and j that are neighbors of each other (relatively to the neighborhood system). The set of all of pairwise interactions is denoted by $\mathcal{W}$ and we set $|\mathcal{W}| = M$.

In this paper we consider first order Discrete Total Variation (DTV) energies [3,6,11] with separable quadratic fidelity. In order to define first order DTVs, let us first follow [21] by noting that a discrete gradient, denoted by ∇ , is defined on any lattice endowed with a neighborhood system. This discrete gradient measures the variation for any two pixels in interaction and is thus a linear mapping $\nabla : \mathbb{R}^N \rightarrow \mathbb{R}^M$. In the remainder of this paper we should consider the matrix representation of ∇ . The adjoint operator of the discrete gradient corresponds to the divergence denoted by div and is defined as usual as

$$\langle \mathbf{u}, \text{div } \mathbf{w} \rangle_N = \langle -\nabla \mathbf{u}, \mathbf{w} \rangle_M.$$

We refer the reader to [21] for more details and properties of these linear operators. In this paper we define the Discrete Total Variation $J(\mathbf{u})$ for any image $\mathbf{u}$ as the l^1 norm of the discrete gradient and is thus defined as

$$J(\mathbf{u}) = \|\nabla \mathbf{u}\|_{1,N} = \sum_{(i,j) \in \mathcal{W}} R_{i,j}(\mathbf{u}) = \sum_{(i,j) \in \mathcal{W}} |u_j - u_i|, \quad (1)$$

where each function $R_{i,j} : \mathbf{u} \in \mathbb{R}^N \mapsto |u_j - u_i| \in \mathbb{R}$ is associated with every single interaction (i, j) . Note that we have only considered non-weighted DTVs for the sake of clarity. Indeed, all results presented in this paper also holds for weighted versions of DTVs (i.e., each interacting term is weighted with a non-negative coefficient).

Finally, let us introduce the notion of subgradient and subdifferential. Let us consider $F : \mathbb{R}^N \rightarrow \bar{\mathbb{R}}$, a proper convex lower semi-continuous function. The *subdifferential* of F at the point $\mathbf{x}$ corresponds to a set, denoted by $\partial F(\mathbf{x})$, defined as the following:

$$\partial F(\mathbf{x}) = \{\mathbf{s} \in \mathbb{R}^N \mid \forall \mathbf{y} \in \mathbb{R}^N, \langle \mathbf{y} - \mathbf{x}, \mathbf{s} \rangle + F(\mathbf{x}) \leq F(\mathbf{y})\}.$$

When F is differentiable at the point $\mathbf{x}$ then $\partial F(\mathbf{x})$ is reduced to a singleton that contains the gradient of F evaluated at $\mathbf{x}$. Any element of $\partial F(\mathbf{x})$ is called a *subgradient* of F at $\mathbf{x}$. Also, for all $p \in \{1, \dots, N\}$ and $\mathbf{x} \in \mathbb{R}^N$, $\partial_p F(\mathbf{x})$ denoted the subdifferential at $\mathbf{x}$ of the partial function $F_{(x_k)_{k \neq p}} : u \in \mathbb{R} \mapsto F(x_1, \dots, x_{p-1}, u, x_{p+1}, \dots, x_N)$. We refer the reader to [15,16] for a complete study of these notions.

3 Differential Inclusion

This section is devoted to the presentation of the differential inclusion that describes the trajectory of the first process, i.e., the DTV-flow. We first recall some theoretical results that are essential for the theoretical soundness of our approach as well as for the computational point of view.

We consider the following differential inclusion problem that formally writes as

$$\begin{cases} \frac{d\mathbf{u}}{dt}(t) \in -\partial J(\mathbf{u}(t)) & \text{on } (0, +\infty) \\ \mathbf{u}(0) = \mathbf{v} \end{cases}, \quad (2)$$

Note that the non-differentiability of J may lead to several possible solutions. Following [2], we should consider the slowest solution that essentially consists in considering the subgradient of J at the current point with minimal Euclidean norm instead of the full subdifferential. Such a point of view not only yields a solution of the differential inclusion but also reduces it to a differential equation. We now list some fundamental properties that are proved for instance in [2].

Proposition 1. *Let $F : \mathbb{R}^N \rightarrow \bar{\mathbb{R}}$ be a proper lower semicontinuous convex function. The differential inclusion (2) has a unique absolutely continuous solution $\mathbf{u}(\cdot) : [0, +\infty) \rightarrow \mathbb{R}$ that satisfies*

$$\frac{d\mathbf{u}}{dt}(t) = -m(\partial F(\mathbf{u}(t))) \text{ for all } t \geq 0, \quad (3)$$

where $m(\partial F(\mathbf{u}(t)))$ is the orthogonal projection of the origin onto $\partial F(\mathbf{u}(t))$. Besides, the following holds:

- (i) $t \mapsto \|\mathbf{u}'(t)\|$ is nonincreasing.
- (ii) $t \mapsto \mathbf{u}(t)$ converges toward a minimizer of F .

In this paper we are interested in the trajectory generated when F corresponds to a discrete Total Variation given by (1) and where the initial condition is set to the observed data. This process generates a trajectory which simplifies the image and eventually converges toward a constant image. This yields a Discrete Total Variation flow that has been used in image processing (see for instance [14]).

Let us just note that the whole trajectory using DTV can be exactly computed using a network flow approach [21]. Such an approach relies on the fact that the trajectory $t \mapsto \mathbf{u}(t)$ is piecewise affine (one way to show this piecewise affine property is to reformulate DTV such that it fits the form of equation (3.4.7) of [15, p. 380]) We refer the reader to [9] for more details on the computations of such a trajectory.

4 A Discrete Total Variation Scale Space Approach

This section describes our approach. First, we show how to preserve the relative order of an image as a variational formulation using Bregman distance. Then, we describe how the latter is coupled with the differential inclusion to yield a contrast preserving denoising approach. Some numerical results are eventually presented.

4.1 Relative Order and Bregman Distances

The goal of this subsection is to characterize the set of all images which have the same *relative order* with respect to a given image using generalized Bregman distances.

Two images $\mathbf{u}$ and $\mathbf{v}$ are said to have the same relative order if and only if the following assertion holds: For all interactions $(i, j) \in \mathcal{W}$ we have

$$(\text{If } v_i < (>)v_j \text{ then } u_i \leq (>)u_j) \text{ or } (\text{If } v_i = v_j \text{ then } u_i = u_j) . \quad (4)$$

Following [20], we introduce the generalized Bregman distance which is a multi-valued version of the standard Bregman distance (which is originally defined for differentiable convex functions) based on the notion of subgradients. More precisely, the generalized Bregman distance $D_G(\mathbf{u}, \mathbf{v})$ between $\mathbf{u}$ and $\mathbf{v}$, relatively to the lower semi-continuous function $G : \mathbb{R}^N \rightarrow \bar{\mathbb{R}}$, is the set of all quantities defined as the following:

$$D_G^{\mathbf{p}(\mathbf{v})}(\mathbf{u}, \mathbf{v}) = G(\mathbf{u}) - G(\mathbf{v}) - \langle \mathbf{p}(\mathbf{v}) | \mathbf{u} - \mathbf{v} \rangle_N, \quad (5)$$

where $\mathbf{p}(\mathbf{v}) \in \partial G(\mathbf{v})$.

Our goal is to describe the relative order set with the help of generalized Bregman distance for DTV, i.e., with $G = J$. More precisely, we will show that by choosing an adequate particular subdifferential $\mathbf{p}(\mathbf{v})$ (among all possible ones in the set $\partial J(\mathbf{v})$), then any image $\mathbf{u}$ which has the same *relative order* with respect to $\mathbf{v}$ satisfies $D_J^{\mathbf{p}(\mathbf{v})}(\mathbf{u}, \mathbf{v}) = 0$. To determine this subgradient, we use the combinatorial structure of DTV. We show that it is enough to consider the subgradients of the functions $R_{i,j}$ associated to each interaction (i, j) . Besides, we are able to express this set with generalized Bregman distance associated with each function $R_{i,j}$.

Let us detail this step. Consider any function $R_{i,j} : \mathbb{R}^N \rightarrow \mathbb{R}$ associated to the interaction (i, j) and its associated generalized Bregman distance. In order to define the latter (following the definition given by (5)), we need to pick an element of the subgradient of $R_{i,j}$ evaluated at the point $\mathbf{v}$. We consider the subgradient of minimal Euclidean norm living in $\partial R_{i,j}(\mathbf{v})$ that is referred to as $m(\partial R_{i,j}(\mathbf{v}))$. The choice of picking the one with minimal Euclidean norm is essential for our approach and will be justified later. The following proposition which relates DTV generalized Bregman distances and relative order is essential for our approach.

Proposition 2. *Let $\mathbf{v}$ be an image. An image $\mathbf{u}$ has the same relative order as $\mathbf{v}$ if and only if $\mathbf{u}$ verifies the following equalities:*

$$\forall (i, j) \in \mathcal{W}, \quad D_{R_{i,j}}^{m(\partial R_{i,j}(\mathbf{v}))}(\mathbf{u}, \mathbf{v}) = |u_j - u_i| + m_i(\partial R_{i,j}(\mathbf{v}))(u_j - u_i) = 0, \quad (6)$$

where $m_i(\partial R_{i,j}(\mathbf{v}))$ is the value of the image $m(\partial R_{i,j}(\mathbf{v}))$ at the site i , i.e., $m_i(\partial R_{i,j}(\mathbf{v})) = (m(\partial R_{i,j}(\mathbf{v})))_i$.

The proof of this proposition is given in Appendix A. Let us note that the proposition may fail to hold for if some particular choice of subgradients if it is not the one of minimal Euclidean norm. Indeed, equality (6) rewrites for any $\mathbf{p}(\mathbf{v}) \in \partial(R_{i,j}(\mathbf{v}))$ and $\mathbf{u} \in \mathbb{R}^N$ as the following (see Appendix A):

$$D_{R_{i,j}}^{\mathbf{p}(\mathbf{v})}(\mathbf{u}, \mathbf{v}) = |u_j - u_i| + p_i(\mathbf{v})(u_j - u_i) = 0, \quad (7)$$

where $p_i(\mathbf{v}) = (\mathbf{p}(\mathbf{v}))_i$ (i.e., the value of the image $\mathbf{p}(\mathbf{v})$ at the site i). Now let us consider only interactions for which we have $v_i = v_j$. For those interactions we have that $p_i(\mathbf{v}) \in [-1, 1]$. We shall see that equality (7) is not necessarily equivalent to the relative order property, depending on the element picked in the subgradient:

1. If $p_i(\mathbf{v}) = 1$, then, according to (7) we have : $D_{R_{i,j}}^{\mathbf{p}(\mathbf{v})}(\mathbf{u}, \mathbf{v}) = 0 \Leftrightarrow u_i \geq u_j$. Thus, there are some images satisfying equality constraint (7) which do not satisfy the relative order of image $\mathbf{v}$ for the interaction (i, j) .

2. If $p_i(\mathbf{v}) = -1$, then, according to (7) we have : $D_{R_{i,j}}^{\mathbf{p}(\mathbf{v})}(\mathbf{u}, \mathbf{v}) = 0 \Leftrightarrow u_i \leq u_j$. As for the previous case, the equivalence may not hold.
3. If $p_i(\mathbf{v}) \in]-1, 1[$, then according to (7) we have : $D_{R_{i,j}}^{\mathbf{p}(\mathbf{v})}(\mathbf{u}, \mathbf{v}) = 0 \Leftrightarrow u_i = u_j$. This mean that all images satisfying equality constraint (7) follows the relative order of the image $\mathbf{v}$ for the interaction (i, j) .

For cases where $v_i \neq v_j$ the subdifferential is reduced to one element and it is easy to check that the equivalence holds. Therefore, instead of picking the subgradient with minimal norm we could have chosen any one for which the i^{th} component is different from 1 or -1 when $v_i = v_j$ (and the pick the only possible subgradient for the other cases). We choose the one with minimal norm not only because it makes the equivalence to hold but also because it is essential for the convergence and thus soundness of our approach.

4.2 A Coupled Scale-Space Approach

We are now ready for describing our approach.

First we consider the unique slow solution of the differential inclusion (2) with discrete Total Variations, i.e., $F = J$ with the observed image $\mathbf{v}$ as the initial condition. Let us note this solution as $t \mapsto u(t)$, and we consider the sequence $(\mathbf{u}(t))_{t \geq 0}$. As the solution corresponds to a Discrete Total Variation flow, we have that a scale space is generated such that the original image is more and more simplified as time evolves. Although one can only use DTV-flow for denoising such as in [13] for instance, it also suffers a loss of contrast [17] as for denoising by minimizing Total Variation with quadratic data fidelity term [23]. Thus, we wish to recover the contrast while keeping the important features contained in the flow.

For this purpose, we introduce a second sequence $(\tilde{\mathbf{u}}(t))_{t \geq 0}$ that are obtained from $(\mathbf{u}(t))_{t \geq 0}$. The most important features we wish to preserve in this paper is the shape of the level so that no geometric information will be introduced. Besides, we also wish to maintain the relative order of gray-level value between two adjacent pixels. Such constraints correspond to maintaining a *relative order*. In order to recover the contrast we wish to search for the image that is the closest (in the sense of the Euclidean norm) to the observed one while satisfying the relative order.

More precisely, we generate the sequence $(\tilde{\mathbf{u}}(t))_{t \geq 0}$ from $(\mathbf{u}(t))_{t \geq 0}$ such that for all $t \in \mathbb{R}^+$ the image $\tilde{\mathbf{u}}(t)$ is the unique solution of the following quadratic convex constrained problem

$$\tilde{\mathbf{u}}(t) := \arg \min_{\mathbf{g} \in S(t)} \left\{ \frac{1}{2} \|\mathbf{g} - \mathbf{v}\|_2^2 \right\}, \quad (8)$$

which corresponds to the projection of observed image $\mathbf{v}$ into closed convex set $S(t)$. This set corresponds to images whose relative orders the same as $\mathbf{u}(t)$. It is formally defined as the following:

1. Set $\mathbf{u}(0) = \tilde{\mathbf{u}}(0) = \mathbf{v}$
2. Solve the differential inclusion problem

$$\begin{cases} \mathbf{u}'(t) \in -\partial J(\mathbf{u}(t)) & \text{on } (0, +\infty) \\ \mathbf{u}(0) = \mathbf{v} \end{cases},$$

3. For all $(i, j) \in \mathcal{W}$, compute $m_i(\partial R_{i,j})$ as below

$$m_i(\partial R_{i,j}(\mathbf{u}(t))) = \begin{cases} -1 & \text{if } \mathbf{u}(t)_i > \mathbf{u}(t)_j, \\ 0 & \text{if } \mathbf{u}(t)_i = \mathbf{u}(t)_j, \\ 1 & \text{if } \mathbf{u}(t)_i < \mathbf{u}(t)_j. \end{cases}$$

4. Solve

$$\tilde{\mathbf{u}}(t) := \arg \min_{\mathbf{g} \in S(t)} \left\{ \frac{1}{2} \|\mathbf{g} - \mathbf{v}\|_2^2 \right\}$$

Fig. 1. Our procedure

$$\begin{aligned} S(t) &= \bigcap_{(i,j) \in \mathcal{W}} \left\{ \mathbf{g} \in \mathbb{R}^N \mid D_{R_{i,j}}^{m(\partial R_{i,j}(\mathbf{u}(t)))}(\mathbf{g}, \mathbf{u}(t)) = 0 \right\} \\ &= \bigcap_{(i,j) \in \mathcal{W}} \left\{ \mathbf{g} \in \mathbb{R}^N \mid |g_j - g_i| + m_i(\partial R_{i,j}(\mathbf{u}(t)))(g_j - g_i) = 0 \right\}, \end{aligned} \quad (9)$$

where $m_i(\partial R_{i,j}(\mathbf{u}(t)))$ is given by

$$m_i(\partial R_{i,j}(\mathbf{u}(t))) = \begin{cases} -1 & \text{if } \mathbf{u}(t)_i > \mathbf{u}(t)_j, \\ 0 & \text{if } \mathbf{u}(t)_i = \mathbf{u}(t)_j, \\ 1 & \text{if } \mathbf{u}(t)_i < \mathbf{u}(t)_j. \end{cases}$$

Our approach is summarized in Figure 1. We can further give a result of convergence for this algorithm which is summarized in the following proposition.

Proposition 3. *Assume that the observed image $\mathbf{v}$ has zero mean, then the sequence of images $(\tilde{\mathbf{u}}(t))_{t>0}$ generated by previous theoretical algorithm converges toward the constant image $\mathbf{0}$.*

We only give the fundamental elements of the proof. According to theorem 1 of [2, p.147], theorem 1 of [2, p.149] and theorem 2 of [2, p.160], we prove that $\lim_n m(\partial R_{i,j}(\mathbf{u}(t))) = \mathbf{0}$ and also that the sequence $(\tilde{\mathbf{u}}(t))_{t>0}$ is bounded. According to theorem 4.18 [22, p.120] and proposition 4.9 [22, p.120] we conclude that $\mathbf{0}$ is unique cluster point of sequence $(\tilde{\mathbf{u}}(t))_{t>0}$.

Proposition 3 yields a natural stopping rule. Assuming the variance σ^2 of the noise that corrupts the image is known, one natural stopping criteria is to stop the process as soon as the residual satisfies $\|\tilde{\mathbf{u}}(t) - \mathbf{v}\|_{2,N}^2 \geq N\sigma^2$. Note that if $\|\mathbf{v}\|_{2,N}^2 > N\sigma^2$ then the stopping criteria is never met and thus the denoised image is the null image (this behavior is similar to the existence of Lagrange

multiplier for Total Variation minimization with quadratic constraint [7]). Also, although the function $t \mapsto \mathbf{u}(t)$ is a continuous function, $t \mapsto \tilde{\mathbf{u}}(t)$ might not be (actually it is not for practical interesting cases).

4.3 Numerical Results

From a practical point of view we take benefit from the fact that the solution of the differential inclusion with Discrete Total Variation given in (2) is piecewise affine. It is thus enough to detect specific times where the subgradient of J with minimal norm is changing to be able to generate the trajectory. This can be easily computed using a network flow approach (see [9]). On a time interval on which the solution of the differential inclusion is affine, its subgradient of minimal Euclidean norm is constant and thus the convex constraint set $S(\cdot)$ is constant of this interval, and thus the solution $\tilde{\mathbf{u}}(\cdot)$ is constant on this interval. It can further be shown that there is a finite number of specific time for which the subgradient of minimal Euclidean norm is changing. Thus, the implementation detects these specific time, compute the solution $\tilde{\mathbf{u}}$ at this time. It iterates until the stopping criteria is met (or the subgradient of minimal Euclidean norm is zero).

In this experiment, we consider Discrete Total Variations that involves the 8-connectivity. The weights for the 4 nearest neighbors are set to 1, while the weights for the diagonal ones are set to $1/\sqrt{2}$.

Figure 2 depicts an original image and its noisy version corrupted by an independent Gaussian additive noise of zero mean and variance $\sigma^2 = 12^2$. The result using our approach is depicted in Figure 3 along with its residual. One can see that the residual seems to contain very little geometric information. The mean square error (MSE) is 48.6191.

(a) original image

(b) noisy image

Fig. 2. Original image is depicted in (a) while its corrupted version by an additive Gaussian noise (zero mean) of variance $\sigma^2 = 12^2$ is depicted in (b)

(a) Our result

(b) Residual

Fig. 3. Result using our approach: The denoised image and and its difference with the noisy image are respectively depicted in (a) and (b)

(a) TV minimizer

(b) Residual

Fig. 4. Total Variation result: The minimizer and and its difference with the noisy image are respectively depicted in (a) and (b)

For comparison puposes we also present in Figure 4 the result obtained using a standard variational model involving Discrete Total Variation with a quadratic constraint ; we refer the reader to [7,23] more details. The MSE for this approach is 49.26. Although the MSE for both approaches are essentially the same it is clearly seen that the residual is much better for approach as it contains much less geometrical detail.

5 Conclusion

This paper has described an approach to restore images while preserving the contrast. It consists on coupling the solution a simplification process, here a DTV-flow, with another process that recovers the contrast. We have shown the well-soundness and the convergence of our approach.

Further extension consists in replacing the Total Variation flow by more elaborated flows. Also, although our approach is a forward scale-space it would be interesting to better understand the theoretical links, if any, with the *inverse* scale space of Osher et al. [20].

Acknowledgement. Research of Jérôme Darbon has been supported by ONR N000140710810. Jérôme Darbon wishes to thank Gary Hower (NAVAIR Weapons Division) for fruitful discussions and his support.

References

1. Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing. Springer, Heidelberg (2002)
2. Aubin, J.P., Cellina, A.: Differential Inclusions. Springer, Heidelberg (1984)
3. Bouman, C., Sauer, K.: A generalized gaussian image model for edge-preserving map estimation. *IEEE Transactions on Signal Processing* 2(3), 296–310 (1993)
4. Burger, M., Gilboa, G., Osher, S., Xu, J.: Nonlinear inverse scale space methods. *Communications in Mathematical Sciences* 5(1), 175–208 (2006)
5. Chalmond, B.: Modeling and Inverse Problems in Image Analysis. Springer, Heidelberg (2003)
6. Chambolle, A., Darbon, J.: On total variation minimization and surface evolution using parametric maximum flows. *International Journal of Computer Vision* 84(3), 288–307 (2009)
7. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. *Num. Math.* (76), 167–188 (1997)
8. Charbonnier, P., Blanc-Féraud, L., Barlaud, M., Aubert, G.: Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing* 5(2), 298–311 (1997)
9. Darbon, J.: In preparation. Tech. rep. (2010)
10. Darbon, J., Ciril, I., Marquina, A., Chan, T., Osher, S.: A note on the bregmanized total variation and dual forms. In: 16th IEEE International Conference on Image Processing (ICIP), pp. 2965–2968 (November 2009)
11. Darbon, J., Sigelle, M.: Image restoration with discrete constrained Total Variation part I: Fast and exact optimization. *Journal of Mathematical Imaging and Vision* 26(3), 261–276 (2006)
12. Geman, D., Reynolds, G.: Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(3), 367–383 (1992)
13. Gilboa, G., Darbon, J., Osher, S., Chan, T.F.: Nonlocal convex functionals for image regularization. Tech. Rep. CAM-06-57, UCLA (October 2006)

14. Gilboa, G., Sochen, N., Zeevi, Y.Y.: Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE Transactions on Image Processing* 11(2), 689–703 (2002)
15. Hiriart-Urruty, J.B., Lemaréchal, C.: *Convex Analysis and Minimization Algorithms Part I*. Springer, Heidelberg (1996)
16. Hiriart-Urruty, J.B., Lemaréchal, C.: *Convex Analysis and Minimization Algorithms Part II*. Springer, Heidelberg (1996)
17. Meyer, Y.: *Oscillating patterns in image processing and nonlinear evolution equations*. University Lecture Series, vol. 22. American Mathematical Society, Providence (2001); the fifteenth Dean Jacqueline B. Lewis memorial lectures
18. Nikolova, M., Chan, R.H.: The equivalence of half-quadratic minimization and the gradient linearization iteration. *IEEE Transactions on Image Processing* 16(6), 1623–1627 (2007)
19. Nikolova, M., Ng, M.: Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific Computing* 27(3), 937–9660 (2005)
20. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An Iterative Regularization Method for Total Variation Based Image Restoration. *SIAM Journal on Multiscale Modeling and Application* 4, 460–489 (2005)
21. Rockafellar, R.T.: *Network Flows and Monotropic Optimization*. John Wiley & Sons, Chichester (1984)
22. Rockafellar, R.T., Wets, R.: *Variational Analysis*. Springer, Heidelberg (1998)
23. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* 60, 259–268 (1992)
24. Strong, D., Chan, T.F.: Edge-Preserving and Scale-Dependent Properties of Total Variation Regularization. *Inverse Problems* 9, S165–S187 (2000)
25. Winkler, G.: *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Applications of Mathematics. Springer, Heidelberg (2003)

A Appendix

Proof of proposition 2 : Let $\mathbf{v} \in \mathbb{R}^N$, be an image and consider an interaction $(i, j) \in \mathcal{W}$. First, for all $\mathbf{p}(\mathbf{v}) \in \partial R_{i,j}(\mathbf{v})$ and $\mathbf{u} \in \mathbb{R}^N$, according to proposition 11.3 [22] we can rewrite $D_{R_{i,j}}^{\mathbf{p}(\mathbf{v})}(\mathbf{u}, \mathbf{v})$ defined by (5) as below :

$$D_{R_{i,j}}^{\mathbf{p}(\mathbf{v})}(\mathbf{u}, \mathbf{v}) = R_{i,j}(\mathbf{u}) + R_{i,j}^*(\mathbf{p}(\mathbf{v})) - \langle \mathbf{p}(\mathbf{v}) | \mathbf{u} \rangle_N, \quad (10)$$

where $R_{i,j}(\mathbf{u}) = r(\mathbf{d}_{i,j}^t \mathbf{u})$, with $r : x \in \mathbb{R} \mapsto |x|$ and $\mathbf{d}_{i,j} \in \mathbb{R}^N$ defined such that: $(\mathbf{d}_{i,j})_i = -1$, $(\mathbf{d}_{i,j})_j = 1$ and for all $p \in \{1, \dots, N\} - \{i, j\}$ $(\mathbf{d}_{i,j})_p = 0$. Also, by invoking [22] theorem 11.23(b) we have :

$$R_{i,j}^*(\mathbf{p}(\mathbf{v})) = \inf_{w \in \mathbb{R}} \{ r^*(w) \mid -w \mathbf{d}_{i,j} = \mathbf{p}(\mathbf{v}) \}, \quad (11)$$

where :

$$r^*(x) = \begin{cases} 0 & \text{if } |x| \leq 1 \\ +\infty & \text{otherwise} \end{cases}. \quad (12)$$

As $\|\mathbf{p}(\mathbf{v})\|_\infty \leq 1$, and according to (11) and (12), the generalized Bregman distance (10) becomes:

$$\begin{aligned} D_{R_{i,j}}^{\mathbf{p}(\mathbf{v})}(\mathbf{u}, \mathbf{v}) &= \inf_{w \in \mathbb{R}} \{ |u_j - u_i| + r^*(w) + w \langle \mathbf{d}_{i,j} | \mathbf{u} \rangle_N - w \mathbf{d}_{i,j} = \mathbf{p}(\mathbf{v}) \} \\ &= |u_j - u_i| + p_i(\mathbf{v})(u_j - u_i) . \end{aligned} \quad (13)$$

By a simple subdifferential calculus we have:

$$\partial_i R_{i,j}(\mathbf{v}) = \begin{cases} -1 & \text{if } v_i > v_j \\ [-1, 1] & \text{if } v_i = v_j \\ 1 & \text{if } v_i < v_j \end{cases} \Rightarrow m_i(\partial R_{i,j}(\mathbf{v})) = \begin{cases} -1 & \text{if } v_i > v_j \\ 0 & \text{if } v_i = v_j \\ 1 & \text{if } v_i < v_j \end{cases} \quad (14)$$

Therefore, taking $\mathbf{p}(\mathbf{v}) = m(\partial R_{i,j}(\mathbf{v}))$ in (13), and according to (14), we have the following three cases:

1. If $v_i > v_j$, we have the following equivalences:

$$D_{R_{i,j}}^{m(\partial R_{i,j}(\mathbf{v}))}(\mathbf{u}, \mathbf{v}) = |u_j - u_i| + (u_j - u_i) = 0 \Leftrightarrow u_i - u_j = |u_j - u_i| \Leftrightarrow u_i \geq u_j .$$

2. If $v_i < v_j$, we have the following equivalences:

$$D_{R_{i,j}}^{m(\partial R_{i,j}(\mathbf{v}))}(\mathbf{u}, \mathbf{v}) = |u_j - u_i| - (u_j - u_i) = 0 \Leftrightarrow u_j - u_i = |u_j - u_i| \Leftrightarrow u_j \geq u_i .$$

3. If $v_i = v_j$, we have the following equivalence:

$$D_{R_{i,j}}^{m(\partial R_{i,j}(\mathbf{v}))}(\mathbf{u}, \mathbf{v}) = |u_j - u_i| = 0 \Leftrightarrow u_i = u_j .$$

Therefore, the equivalence stated in proposition 2 is proved. $\square$