A NARROW BAND METHOD FOR THE CONVEX FORMULATION OF DISCRETE MULTI-LABEL PROBLEMS

ANTONIO BAEZA *, VICENT CASELLES †, PAU GARGALLO ‡, AND NICOLAS PAPADAKIS §

Abstract. We study a narrow band type algorithm to solve a discrete formulation of the convex relaxation of energy functionals with total variation regularization and non convex data terms. We prove that this algorithm converges to a local minimum of the original non linear optimization problem. We illustrate the algorithm with experiments for disparity computation in stereo and a multi-label segmentation problem and we check experimentally that the energy of the local minimum is very near to the energy of the global minimum obtained without the narrow band type method.

Key words. convex relaxation, anisotropic total variation, multilabel problems.

AMS subject classifications. 52A38, 68U10, 65K10

1. Introduction.

The purpose of this paper is to describe a narrow band type method to minimize the convex relaxation of energy functionals with total variation regularization and non convex data terms as proposed in [20, 14].

While binary labeling problems can be globally minimized using graph cuts [16, 18], in general, the global optimum of multilabel problems cannot be obtained, although it can be solved approximately [5, 23]. As Ishikawa showed in [17], in case that the set of labels is linearly ordered and the pairwise interactions between them are convex, then the multilabel problem can be exactly solved. Assuming that the labeling function u is defined on a discrete grid, the main idea was to transform this problem into a binary labeling problem for the characteristic function of the subgraph of u. This idea was reformulated in [20] in order to construct a convex relaxation of the continuous variational formulation of total variation regularized problems with non convex data terms. The applications to image processing are numerous, in particular, to stereo, optical flow computation, or multi-region image segmentation. One of the advantages of the continuous formulation is that it permits to use numerical optimization techniques which can be implemented in parallel architectures such as GPUs [21].

This set of ideas has been deeply studied and extended in [14] (see also [8]), where the authors addressed in detail the problem of the computation of minimal partitions, with a regularization term given by the Hausdorff measure of the total interface in the continuous setting, or the Potts model in the discrete case, although the methods apply to other cases, like the disparity computation in stereo. A related approach for depth computation has been proposed in [24] and in [19].

If the original unknown function is $u : \{1, \ldots, N\}^2 \to [0, M]$, the new variable in [20] is the characteristic function of the subgraph of u, that is $w(i, j, z) = \chi_{[0,u(i,j)]}(z)$, $(i, j) \in \{1, \ldots, N\}^2$, $z \in [0, M]$. At the discrete level we use a finite set of values

*Barcelona Media, Avenida Diagonal 177, 08018 Barcelona, Spain, antonio.baeza@barcelonamedia.org

†DTIC, Universitat Pompeu Fabra, Roc Boronat, 138, 08018 Barcelona, Spain, vicent.caselles@upf.edu

‡Barcelona Media, Avenida Diagonal 177, 08018 Barcelona, Spain, pau.gargallo@barcelonamedia.org

§Barcelona Media, Avenida Diagonal 177, 08018 Barcelona, Spain, nicolas.papadakis@barcelonamedia.org

for z, for instance $\{0,\ldots,M\}$, assuming that M is an integer. In [14] the authors represent the function in terms of the characteristic functions of its level sets: $u_l(i,j) = \chi_{[0,u(i,j)]}(l)$, $l = 1,\ldots,M$ and they write the energy in terms of $u_1,\ldots,u_M$ which satisfy the constraint $0 \le u_M \le \ldots \le u_1 \le 1$. Thus, at the numerical level we always work with a finite set of values.

Since we have gone from 2D data to 3D data, we have increased both the computational complexity and the memory storage required. The first issue may be not dramatic since the task of energy minimization can be very effective using GPUs, although it will be more efficient if we work in a narrow band around the graph of the current solution. Working in a narrow band, the memory storage will be alleviated at equal number of labels. These are the supporting reasons to study a narrow band type method to minimize the convex relaxation of energy functionals with total variation regularization and non convex data terms.

Let us briefly summarize in a continuous framework the relaxation technique proposed in [20] (see also [14]). Let us consider the variational problem

$$\min_{u\in BV(\Omega),0\le u\le M} \mathcal{R}(u) := \int_\Omega |Du| + \int_\Omega W(x,u(x))\,dx, \tag{1.1}$$

where Ω is a rectangle in $I\!R^2$, $\int_\Omega |Du|$ denotes the total variation of u in Ω [4], and $W : \Omega \times I\!R \to I\!R^+$ is a potential which is Borel measurable in x and continuous in u, but not necessarily convex. Thus, the functional is nonlinear, and let us assume that is non-convex. The functional can be relaxed to a convex one by considering the subgraph of u as an unknown.

Following [20], let us write the nonlinearities in (1.1) in a "convex way" by introducing a new auxiliary variable representing the subgraph of u. This will permit to use standard optimization algorithms. The treatment here will be heuristic.

Let $w(x,s) = H(u(x)-s)$, where $H = \chi_{[0,+\infty)}$ is the Heaviside function and $s \in I\!R$. Notice that the set of points where $u(x) > s$ (the subgraph of u) is identified as $w(x,s) = 1$. That is, $w(x,s)$ is an embedding function for the subgraph of u. This permits to consider the problem as a binary labeling problem. The graph of u is a 'cut' in w.

Since $D_x w(x,s) = \delta(u(x)-s)Du(x)$ and $\partial_s w(x,s) = -\delta(u(x)-s)$ we may formally write $\mathcal{R}(u)$ as

$$\begin{aligned}\mathcal{R}(u) = &\int_\Omega \int_0^M \left(|D_x w| + W(x,s)|\partial_s w(x,s)|\right)\,dx\,ds \\ &+ \int_\Omega \left(W(x,0)|w(x,0)-1| + W(x,1)|w(x,1)|\right)dx := \mathcal{E}(w).\end{aligned}$$

Let

$$\mathcal{A} := \{w \in BV(\Omega\times[0,M]) : w(x,s) \in \{0,1\}, \forall (x,s) \in \Omega\times[0,M],\ \partial_s w \le 0\}.$$

Notice that

$$\mathcal{A} = \{\chi_F : F \text{ is the subgraph of some } u \in BV(\Omega),\quad 0 \le u \le M\ \}.$$

Using the definition of anisotropic total variation [3] we may write the problem (1.1) in terms of w as

$$\min_{w\in\mathcal{A}} \mathcal{E}(w), \tag{1.2}$$

where the boundary conditions $w(x,0) = 1$, $w(x,M) = 0$ are taken in a variational sense.

Although the energy (1.2) is convex in w the problem is non-convex since the minimization is carried on $\mathcal{A}$ which is a non-convex set. The proposal in [20] is to relax the variational problem by allowing w to take values in $[0,1]$. This leads to the following class of admissible functions

$$\mathcal{A}_c := \{w \in BV(\Omega \times [0,M]) : w(x,s) \in [0,1],\ \forall (x,s) \in \Omega \times [0,M],\ \partial_s w \leq 0\}. \quad (1.3)$$

The associated variational problem is written as

$$\min_{w \in \mathcal{A}_c} \mathcal{E}(w). \quad (1.4)$$

Let us mention that the solution of problem (1.4) may not be unique, in consistency with the original problem (1.1). On the other hand, the problem is now convex and a solution can be obtained using a proximal point type algorithm [20, 14]

$$\min_{w \in \mathcal{A}_c} \frac{\|w - w^k\|^2}{2\Delta} + \mathcal{E}(w), \quad (1.5)$$

where $\Delta > 0$ and w^k is the current solution, which can be solved using the dual [11] or primal-dual numerical schemes [25]. In this way a global minimum of (1.4) is obtained. Formally, the level sets of a solution of (1.4) give solutions of (1.1). This has been observed several times (see [2, 15, 20, 6] and references therein). In the most general setting, this can be proved using the developments in [3, 10].

Clearly, this relaxation comes at the prize of using an extra dimension (or an ordered set of functions as in [14]). In order to recover the complexity of a 2D problem, we study a proximal point iterative algorithm formulated in a narrow-band around the graph of the current solution working at the discrete level. At each step, by applying a convenient threshold to the previous solution w^k, we obtain a binary function, $T(w^k)$, and we solve

$$\min_{w} \frac{\|w - T(w^k)\|^2}{2\Delta} + \mathcal{E}(w), \quad (1.6)$$

where the functions w are defined in a neighborhood of the jump discontinuity of $T(w^k)$ (representing the graph of the solution we are looking for) and are decreasing in the variable s. Then we prove that by choosing the thresholding operations in a convenient way, the solutions w^k converge to a local minimum w^* (which is the characteristic function of the subgraph of a function) of the discrete version of (1.1). Indeed, the energy of w^* is smaller that the energy of functions whose support is in a neighborhood of the graph represented by w^* and which represent its perturbations. But in practice we have checked that the energy of the solution obtained is very near to the energy of the solution obtained without using the narrow band method which is a global minimum of (1.4) [20]. We will also show that this provides a more efficient method with lower memory requirements. We have implemented at present the algorithm in CPU and we plan its GPU implementation.

We illustrate our method with some experiments on disparity computation in pairs of stereo images and on the convex relaxation of the piecewise constant discrete version of the Mumford-Shah model, and we compare our results with the ones obtained without the narrow band method [20, 14]. We observe that the local minima obtained are very near to the global minimum computed without the narrow band method

Let us finally mention that the relaxation formalism can be extended to more general energy functionals [1, 14]. In particular, in [14] the authors address the problem of convex formulation of multi-label problems with finitely many values including (1.1) and the case of non-convex neighborhood potentials like the Potts model or the truncated total variation. The general framework permits to consider the relaxation in $BV(\Omega)$ of functionals of the form

$$F(u) := \int_\Omega f(x, u(x), \nabla u(x))\, dx \tag{1.7}$$

where $u \in W^{1,1}(\Omega)$ and $f : \Omega \times I\!\!R \times I\!\!R^N \to [0, \infty[$ is a Borel function such that $f(x, z, \xi)$ is a convex function of ξ for any $(x, z) \in \Omega \times I\!\!R^N$ satisfying some coercivity assumption in ξ. Let f^* denote the Legendre-Fenchel conjugate of f with respect to ξ. If

$$\mathcal{K} := \{w = (w^x, w^s) : \Omega \times I\!\!R \to I\!\!R^2 : w \text{ is smooth and } f^*(x, s, w^x(x, s)) \le w^s(x, s)\},$$

then the lower semicontinuous relaxation of F is

$$\mathcal{F}(u) = \sup_{w \in \mathcal{K}} \int_\Omega \int_{I\!\!R} w \cdot D\chi_{\{(x,s):s<u(x)\}}\, ds\, dx.$$

Based on this formula one can use a dual or a primal-dual numerical scheme to minimize $\mathcal{F}(u)$ if one knows how to compute the projection onto the convex set $\mathcal{K}$. We refer to [14] for details.

Let us give the plan of the paper. In Section 2 we recall the definition and basic properties of discrete total variation (TV) functionals. In Section 3 we propose an abstract formulation of the narrow band method for discrete TV functionals. In Section 4 we introduce the discrete gradient and divergence required to formulate problems of type (1.5) in a discrete setting. Then in Section 5 we study the equivalence of the discrete formulations of (1.1) and (1.4). In Section 6 we formulate the dual problem of the discrete energy (1.5) and write the associated Euler-Lagrange equation. In Section 7 we adapt the abstract narrow band method to the concrete case of the discrete energy (1.5) and we obtain the convergence of the numerical scheme. We also observe that the discrete analogous of the constraint $\partial_s w \le 0$ can be imposed by penalization when the penalization parameter is larger than a multiple of the number of labels. This is technically relevant in order to fix from the beginning the thresholds at each step of the proximal point. In Section 8 we present some experiments that illustrate the algorithm and we check that the energy of the local minima obtained is very near to the global minimum computed without the narrow band method. The experiments are shown for two problems: disparity computation in stereo problems and multi-label segmentation.

2. Basic properties of discrete total variation functionals. We denote by X the Euclidian space $I\!\!R^m$, $m \ge 1$, endowed with the Euclidian metric and the scalar product, denoted by $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$, respectively. As in [13] we define a discrete total variation as a convex, nonnegative functional $J : X \to [0, +\infty]$, not identically $+\infty$, satisfying a coarea formula

$$J(w) = \int_{I\!\!R} J(\chi_{\{w \ge t\}})\, dt, \qquad w \in X. \tag{2.1}$$

As stated in [13], it is easy to derive the following properties.

PROPOSITION 2.1. *Let J be a discrete total variation in X. Then*

(i) *J is positively homogeneous, i.e., $J(\lambda w) = \lambda J(w)$ for any $w \in X$ and any $\lambda \geq 0$.*
(ii) *J is invariant by addition of a constant: $J(c\mathbf{1} + w) = J(w)$ for any $w \in X$ and any $c \in \mathbb{R}$, where $\mathbf{1} = (1, 1, \ldots, 1)$ is a constant vector. In particular $J(\mathbf{1}) = 0$.*
(iii) *J is lower semicontinuous*
(iv) *J satisfies*

$$J(w \wedge \alpha) \leq J(w) \quad \text{and} \quad J(w \vee \alpha) \leq J(w), \quad \forall w \in X, \ \forall \alpha \in \mathbb{R}.$$

(v) *J is submodular: for any $w, w' \in \{0,1\}^m$,*

$$J(w \wedge w') + J(w \vee w') \leq J(w) + J(w')$$

This holds also for any $w, w' \in X$.

Let C be a closed convex subset of X and let us consider the problem

$$\min_{w \in C} J(w). \tag{2.2}$$

Using (2.1), for any $w \in X$ the measure of the set of $t \in \mathbb{R}$ for which $J(\chi_{\{w \geq t\}}) \leq J(w)$ is positive and we always find a value of $t \in \mathbb{R}$ such that $J(\chi_{\{w \geq t\}}) \leq J(w)$. In particular, if w is a minimum of (2.2), then the level sets are also minima of (2.2) as soon as the level sets are also in C.

We assume that $C \subseteq \{w : w_i \in [0,1], \ \forall i \in \{1, \ldots, m\}\}$ satisfies the property:

$$w \in C \quad \text{if and only if} \quad \chi_{\{w \geq t\}} \in C, \ \forall t \in [0,1]. \tag{2.3}$$

We also assume that there is some $w \in C$ such that $J(w) < \infty$.

Observe that property (2.3) implies that $w \wedge \alpha, w \vee \alpha \in C$ for any $w \in C$.

Let $\Delta > 0$ and $g \in X$. Let us consider the problem

$$\min_{w \in C} \frac{1}{2\Delta} \|w - g\|^2 + J(w). \tag{2.4}$$

Since J is convex and $J(w) < \infty$ for some $w \in C$, problem (2.4) has a unique solution $\overline{w} \in C$. Moreover, using Proposition 2.1.(iv), we know that $\min g \leq \overline{w}_i \leq \max g$ for all $i \in \{1, \ldots, m\}$. It suffices to check that $\overline{w} \vee (\min g)$ and $\overline{w} \wedge (\max g)$ are both solutions of (2.4) and must coincide with $\overline{w}$ (see Proposition 7.2).

Let $\mathcal{S}(C) := \{F \subseteq \{1, \ldots, m\} : \chi_F \in C\}$.

Proposition 2.2. *Let $\overline{w} \in C$ be the unique solution of (2.4). Then for any $t \in \mathbb{R}$ the set $\{\overline{w} > t\} := \{i : \overline{w}_i > t\}$ is a solution of*

$$(P)_t \qquad \min_{F \in \mathcal{S}(C)} J(\chi_F) + \frac{1}{\Delta} \langle \chi_F, (t - g) \rangle. \tag{2.5}$$

Moreover, $\{w > t\}$ and $\{w \geq t\}$ are, respectively, the minimal and maximal solutions of $(P)_t$.

Before proving Proposition 2.2 let us recall the following Lemma.

Lemma 2.3. (i) *For each $t \in \mathbb{R}$, the problem $(P)_t$ has a solution.*

(ii) *Let $t > t'$ and let Q_t, $Q_{t'}$ be solutions of $(P)_t$ and $(P)_{t'}$, respectively. Then $Q_t \subseteq Q_{t'}$.*

(iii) If $t \leq a := \min g$, then $\{1, \ldots, m\}$ is a solution of $(P)_t$, unique if $t < a$. If $t \geq b := \max g$, then $\emptyset$ is a solution of $(P)_t$, unique if $t > b$.

The first assertion is immediate. The second one is a consequence of the submodularity of J and can be found in [2, 12]. The third assertion is also immediate.

Proof. (of Proposition 2.2). Recall that we denote $\overline{w} \in C$ the unique solution of (2.4). Let $Q_t \in \mathcal{S}(C)$ be a solution of $(P)_t$, $t \in I\!R$. Let us define

$$w_i = \sup\{t \in I\!R : i \in Q_t\}.$$

The function w_i is well defined, $w \in X$, and $\{w > t\} = Q_t$ a.e. in t. Since w has finitely many values, this implies that all (different) level sets are in C. Hence $w \in C$ by our assumption (2.3). Notice also that, by Lemma 2.3, $a \leq w_i \leq b$ for any $i \in \{1, \ldots, m\}$. Observe that w is a solution of (2.4). Indeed

$$\begin{aligned} J(w) + \frac{1}{2\Delta}\|w-g\|^2 &= \int_a^b J(\chi_{Q_t})\,dt + \frac{1}{\Delta}\int_a^b \sum_{i=1}^m (\chi_{Q_t})_i(t-g_i)\,dt + \sum_{i=1}^m \frac{(a-g_i)^2}{2\Delta} \\ &\leq \int_a^b J(\chi_{\{\overline{w}\geq t\}})\,dt + \frac{1}{\Delta}\int_a^b \sum_{i=1}^m (\chi_{\{\overline{w}\geq t\}})_i(t-g_i)\,dt + \sum_{i=1}^m \frac{(a-g_i)^2}{2\Delta} \\ &= J(\overline{w}) + \frac{1}{2\Delta}\|\overline{w}-g\|^2. \end{aligned}$$

Thus w is also a solution of (2.4). By uniqueness $w = \overline{w}$. Hence $\{\overline{w} > t\} = Q_t$ a.e. in t. We deduce that for any $t \in I\!R$ the set $\{\overline{w} > t\}$ is a solution of $(P)_t$. By Lemma 2.3.*(ii)* it follows that $\{w > t\}$ and $\{w \geq t\}$ are, respectively, the minimal and maximal solutions of $(P)_t$. □

It may happen that the functional $J(\chi_E)$ is not symmetric under complements, i.e., we do not have $J(\chi_{\{1,\ldots,m\}\setminus E}) = J(\chi_E)$. For that reason we define

$$J^*(\chi_E) = J(\chi_{\{1,\ldots,m\}\setminus E})$$

for any set E such that $\{1, \ldots, m\} \setminus E \in \mathcal{S}(C)$.

Proposition 2.4. *Let $\overline{w} \in C$ be the unique solution of (2.4). Then for any $t \in I\!R$ the set $\{\overline{w} < t\} := \{i : \overline{w}_i < t\}$ is a solution of*

$$(P)_t^* \qquad \min_{F\subseteq\{1,\ldots,m\}:F^c\in\mathcal{S}(C)} J^*(\chi_F) - \frac{1}{\Delta}\langle \chi_F, (t-g)\rangle, \tag{2.6}$$

where $F^c := \{1, \ldots, m\} \setminus F$. Moreover, $\{w < t\}$ and $\{w \leq t\}$ are, respectively, the minimal and maximal solutions of $(P)_t^*$.

Proof. Let $F_t := \{w < t\}$, $t \in I\!R$, $F \subseteq \{1, \ldots, m\}$ such that $F^c \in \mathcal{S}(C)$. Then $(F_t)^c = \{w \geq t\} := E_t$ and we have

$$\begin{aligned} J^*(\chi_{F_t}) - \frac{1}{\Delta}\langle \chi_{F_t}, (t-g)\rangle &= J(\chi_{E_t}) + \frac{1}{\Delta}\langle \chi_{E_t}, (t-g)\rangle - \frac{1}{\Delta}\langle \mathbf{1}, (t-g)\rangle \\ &\leq J(\chi_{F^c}) + \frac{1}{\Delta}\langle \chi_{F^c}, (t-g)\rangle - \frac{1}{\Delta}\langle \mathbf{1}, (t-g)\rangle \\ &= J^*(\chi_F) - \frac{1}{\Delta}\langle \chi_F, (t-g)\rangle. \end{aligned}$$

□

3. An abstract formulation of the narrow band method for discrete TV functionals. Let us consider the following iterative scheme. Start with $w^0 \in C$ with $J(w^0) < \infty$. For each n, let w^{n+1} be the solution of

$$\min_{w \in C} \frac{1}{2\Delta} \|w - w^n\|^2 + J(w).$$

Hence (see [7]), for each n, w^{n+1} is a solution of

$$\frac{w^{n+1} - w^n}{\Delta} + \partial(J + I_C)(w^{n+1}) \ni 0, \tag{3.1}$$

where $I_C(w) = 0$ if $w \in C$ and $I_C(w) = +\infty$ otherwise. We write the inclusion symbol $\ni$ instead of the equality since $\partial(J + I_C)$ may be multivalued [7]. As it is known, the proximal point algorithm is convergent [22, 9].

Let us now consider the following modification of the scheme which is an abstract formulation for the narrow band method. Start with $w^0 = \overline{w}^0 \in C$ with $J(w^0) < \infty$, $C_0 = C$. For each n, let $C_n \subseteq X$ be a closed convex set satisfying (2.3). Assume that for each n we know $w^n \in C_n$ and we solve the problem

$$\min_{w \in C_n} \frac{1}{2\Delta} \|w - w^n\|^2 + J(w). \tag{3.2}$$

Let $\overline{w}^{n+1}$ be the unique solution of (3.2). Let us denote $T_t(w) = \chi_{\{w>t\}}$, $w \in X$.

Given $\overline{w}^{n+1} \in C_n$ we may find t_{n+1} such that $J(T_{t_{n+1}}(\overline{w}^{n+1})) \le J(\overline{w}^{n+1}) + \epsilon_{n+1}$ where $\sum_{n=1}^{\infty} \epsilon_n < \infty$. For simplicity, we write T_{n+1} instead of $T_{t_{n+1}}$. We define

$$w^{n+1} = T_{n+1}(\overline{w}^{n+1}).$$

Let $C_{n+1} \subseteq X \cap \{0 \le w \le 1\}$ be a closed convex set satisfying property (2.3) such that $w^{n+1} \in C_{n+1}$. Now problem (3.2) is completely determined for next iteration. We assume that the set C_{n+1} is determined by w^{n+1} in a unique way.

Let us observe the following estimate.

LEMMA 3.1. *We have*

$$\frac{1}{\Delta} \sum_{n=0}^{p} \left\| \overline{w}^{n+1} - w^n \right\|^2 + J(\overline{w}^{p+1}) \le J(\overline{w}^0) + \sum_{n=0}^{p} \epsilon_n. \tag{3.3}$$

This estimate implies that $\overline{w}^{n+1} - w^n \to 0$ *in* X *as* $n \to \infty$.

Proof. For each n, $\overline{w}^{n+1}$ is a solution of

$$\frac{\overline{w}^{n+1} - w^n}{\Delta} + v^{n+1} = 0. \tag{3.4}$$

where $v^{n+1} \in \partial J_n(\overline{w}^{n+1})$, $J_n = J + I_{C_n}$.

Now, multiplying (3.4) by $\overline{w}^{n+1} - w^n$, taking scalar products, and adding the identities obtained we have

$$\frac{1}{\Delta} \sum_{n=0}^{p} \left\| \overline{w}^{n+1} - w^n \right\|^2 = \sum_{n=0}^{p} \langle v^{n+1}, w^n - \overline{w}^{n+1} \rangle \le \sum_{n=0}^{p} (J_n(w^n) - J_n(\overline{w}^{n+1}))$$

$$= \sum_{n=0}^{p} (J(w^n) - J(\overline{w}^{n+1})) \le \sum_{n=0}^{p} (J(\overline{w}^n) - J(\overline{w}^{n+1})) + \sum_{n=0}^{p} \epsilon_n.$$

where we used that $J(w^n) \leq J(\overline{w}^n) + \epsilon_n$ in the last inequality. Then (3.3) follows. □

THEOREM 3.2. *For n large enough the solution becomes a constant and binary function. Thus the algorithm converges in finitely many steps to a local minimum of J, i.e. a minimum of J in C_n.*

Proof. Since the set of characteristic functions in a finite grid, though very big, is finite, then there are m, p with $m > p$ such that $w^m = w^p$. Then $\overline{w}^{m+1} = \overline{w}^{p+1}$. This implies that the sequence $\overline{w}^n$ is periodic. Let $k = m - p$. Then $\{\overline{w}^{jk+p+1}\}_j$ is a constant sequence, say $\tilde{w}$. Let $m_j = jk+p$. Notice that C_{m_j} does not depend on j, let us denote it by $\mathcal{B}$. Since $v^{m_j+1} \in \partial J_{m_j}(\overline{w}^{m_j+1})$, by the definition of subdifferential, we have that

$$J(w) - J(\overline{w}^{m_j+1}) \geq \langle v^{m_j+1}, w - \overline{w}^{m_j+1}\rangle \qquad \forall w \in C_{m_j}. \tag{3.5}$$

Since $\overline{w}^{m_j+1} = \tilde{w}$, $v^{m_j+1} = -\frac{1}{\Delta}(\overline{w}^{m_j+1} - w^{m_j}) \to 0$ in X (by Lemma 3.1) and is independent of j, then $v^{m_j+1} = 0$. Hence $w^{m_j} = \overline{w}^{m_j+1} = \tilde{w}$ and we obtain

$$J(w) - J(\tilde{w}) \geq 0 \qquad \forall w \in \mathcal{B}.$$

In other words $\tilde{w}$ is a minimum of (2.2) in $\mathcal{B}$. Moreover the sequence becomes stationary since $w^{m_j+1} = T_{m_j+1}(\overline{w}^{m_j+1}) = w^{m_j} = \tilde{w}$. Therefore, the algorithm converges in a finite number of steps to a binary function. □

We now focus on determining the threshold value t_{n+1} for each iteration. Let $E_t := \{\overline{w}^{n+1} > t\}$, $F_t := \{\overline{w}^{n+1} \leq t\}$, $t \in [0, 1]$.

Since we know by Lemma 3.1 that $\overline{w}^{n+1} - w^n \to 0$ in X as $n \to \infty$ and all norms in a finite dimensional space are equivalent, then, given $\epsilon > 0$, for n large enough $(n \geq n_\epsilon)$ we have that

$$\|\overline{w}^{n+1} - w^n\|_\infty < \epsilon, \tag{3.6}$$

where $\|\cdot\|_\infty$ denotes the maximum norm in X.

LEMMA 3.3. *Assume that $\epsilon \leq \frac{1}{2}$ and let n be such that (3.6) holds. Then the infimum of $\{J(\chi_{E_t}) : t \in [\min \overline{w}^{n+1}, \max \overline{w}^{n+1})\}$ is attained at one of the values $J(\chi_{\{\overline{w}^{n+1} > \min \overline{w}^{n+1}\}})$ or $J(\chi_{\{\overline{w}^{n+1} \geq \max \overline{w}^{n+1}\}})$.*

Proof. By (3.6), if $w_i^n = 1$, then $|\overline{w}_i^{n+1} - 1| < \epsilon$; if $w_i^n = 0$, then $|\overline{w}_i^{n+1}| < \epsilon$. Then $T_t(\overline{w}^{n+1}) = w^n$ for all $t \in [\epsilon, 1-\epsilon]$. In particular:

$$\{\overline{w}^{n+1} > t\} = \{w^n = 1\} \quad \text{and} \quad \{\overline{w}^{n+1} \leq t\} = \{w^n = 0\} \quad \forall t \in [\epsilon, 1-\epsilon].$$

Let $\epsilon \leq t < t' < \max \overline{w}^{n+1}$. Then $E_{t'} \subseteq E_t \subseteq \{w^n = 1\}$. Then by Proposition 2.2 we have

$$J(\chi_{E_{t'}}) + \frac{1}{\Delta}\int_{E_{t'}} (t' - w^n) \leq J(\chi_{E_t}) + \frac{1}{\Delta}\int_{E_t} (t' - w^n).$$

Hence

$$J(\chi_{E_{t'}}) \leq J(\chi_{E_t}) + \frac{1}{\Delta}\int_{E_t \setminus E_{t'}} (t' - w^n) \leq J(\chi_{E_t}).$$

Now, let $\min \overline{w}^{n+1} \leq t < t' \leq 1 - \epsilon$. Then $F_t \subseteq F_{t'} \subseteq \{w^n = 0\}$. Then

$$J^*(\chi_{F_t}) - \frac{1}{\Delta}\int_{F_t} (t - w^n) \leq J^*(\chi_{F_{t'}}) - \frac{1}{\Delta}\int_{F_{t'}} (t - w^n).$$

Hence

$$J^*(\chi_{F_t}) \leq J^*(\chi_{F_{t'}}) - \frac{1}{\Delta}\int_{F_{t'}\setminus F_t}(t-w^n) \leq J^*(\chi_{F_{t'}}).$$

Since $J^*(\chi_E) = J(\chi_{\{1,\ldots,N\}\setminus E})$ we have

$$J(\chi_{E_t}) \leq J(\chi_{E_{t'}}) \qquad \forall \min \overline{w}^{n+1} \leq t < t' \leq 1-\epsilon.$$

In that circumstance the infimum of $J(\chi_{E_t})$ for any $t \in [\min \overline{w}^{n+1}, \max \overline{w}^{n+1})$ is attained one of the values $J(\chi_{\{\overline{w}^{n+1} > \min \overline{w}^{n+1}\}})$ or $J(\chi_{\{\overline{w}^{n+1} \geq \max \overline{w}^{n+1}\}})$. □

For some operators we can use the previous scheme starting at iteration $n = 0$. The proof of the next statement follows the same line as the proof of Lemma 3.3 with $\epsilon = \frac{1}{2}$.

Lemma 3.4. *Assume that* $\kappa := \sup_{w\in X, 0\leq w\leq 1} \|\partial J(w)\|_\infty < \infty$. *Let* $\Delta > 0$ *be such that* $\kappa\Delta < \frac{1}{2}$. *For any* n, *if* $E_t = \{w^{n+1} > t\}$, *then* $J(\chi_{E_t})$ *is decreasing for* $t \in (\frac{1}{2}, \max \overline{w}^{n+1})$ *and increasing for* $t \in (\min \overline{w}^{n+1}, \frac{1}{2})$.

Thus, the infimum of $\{J(\chi_{E_t}) : t \in [\min \overline{w}^{n+1}, \max \overline{w}^{n+1})\}$ is one of the values $J(\chi_{\{\overline{w}^{n+1} > \min \overline{w}^{n+1}\}})$ or $J(\chi_{\{\overline{w}^{n+1} \geq \max \overline{w}^{n+1}\}})$. Then we take the characteristic function of the corresponding level set as w^{n+1}.

Notice that the set $\{\overline{w}^{n+1} > \min \overline{w}^{n+1}\}$ contains the set $\{w^n = 1\}$ while $\{\overline{w}^{n+1} \geq \max \overline{w}^{n+1}\}$ is contained in it.

4. The discrete gradient and divergence for a problem with mixed Neumann and Dirichlet boundary conditions. Let us consider the discrete domain $\Omega = \{1,\ldots,N\}^2$. Let $\Omega_* := \Omega \times \{1,\ldots,M\}$. For convenience, let us denote $\tilde{\Omega}$ the extended domain $\Omega \times \{0,\ldots,M\}$. We denote by U_*, U the Euclidian spaces $\mathbb{R}^{N\times N\times M}$, $\mathbb{R}^{N\times N\times (M+1)}$, respectively.

Let us denote the Euclidian norm and scalar product in U by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$, respectively. The Euclidian norm an scalar product in U_* will be denoted by $\|w\|_*$ and $\langle\cdot,\cdot\rangle_*$, respectively. Thus, we write

$$\|w\|^2 = \sum_{(i,j,k)\in\tilde{\Omega}} |w_{i,j,k}|^2 \qquad w \in U$$

$$\|w\|_*^2 = \sum_{(i,j,k)\in\Omega_*} |w_{i,j,k}|^2 \qquad w \in U_*.$$

Let $\langle\cdot,\cdot\rangle_0$ be the scalar product for the vectors $(w_{i,j,0})_{(i,j)\in\Omega}$.

The definition of the discrete gradient will be in consonance with the boundary conditions. Recall that we need to impose the Neumann boundary conditions when $i = 1$ or $i = N$ and when $j = 1$, $j = N$, and Dirichlet boundary conditions when $k = 0$, $k = M$. For that reason, although the original domain is Ω_* and we are interested in the values of $w_{i,j,k}$ for $(i,j,k) \in \Omega_*$, to impose the Dirichlet boundary conditions we need to consider vectors $w \in U$ with $w_{i,j,0} = 1$. For the zero Dirichlet boundary conditions when $k = M$ we do not write explicitly the extension of the domain to $k = M + 1$.

Let us give the definition of the discrete gradient which considers Dirichlet boundary conditions on the top and bottom of $\tilde{\Omega}$ and Neumann in the lateral boundary of

$\tilde{\Omega}$. Given $w \in U$ its discrete gradient ∇w will be a vector in $V := U \times U \times U$ given by $(\nabla w)_{i,j,k} = ((\nabla_x w)_{i,j,k}, (\nabla_y w)_{i,j,k}, (\nabla_s w)_{i,j,k})$, $(i,j,k) \in \tilde{\Omega}$, where

$$(\nabla_x w)_{i,j,k} = \begin{cases} w_{i+1,j,k} - w_{i,j,k} & \text{if } i < N \\ 0 & \text{if } i = N, \end{cases} \tag{4.1}$$

$$(\nabla_y w)_{i,j,k} = \begin{cases} w_{i,j+1,k} - w_{i,j,k} & \text{if } j < N \\ 0 & \text{if } j = N, \end{cases} \tag{4.2}$$

$$(\nabla_s w)_{i,j,k} = \begin{cases} w_{i,j,k+1} - w_{i,j,k} & \text{if } k < M \\ -w_{i,j,k} & \text{if } k = M. \end{cases} \tag{4.3}$$

Notice that, since $w_{i,j,0} = 1$, we have $(\nabla_s w)_{i,j,0} = w_{i,j,1} - 1$ for all $(i,j) \in \Omega$.

Let us denote $(\nabla_{xy} w)_{i,j,k} = ((\nabla_x w)_{i,j,k}, (\nabla_y w)_{i,j,k})$.

The scalar product and the norm in V are denoted respectively by $\langle \tilde{\sigma}, \tilde{\eta} \rangle_V = \sum_{i,j,k \in \tilde{\Omega}} \tilde{\sigma}_{i,j,k} \cdot \tilde{\eta}_{i,j,k}$ and $\|\tilde{\sigma}\|_V^2 = \langle \tilde{\sigma}, \tilde{\sigma} \rangle_V$.

Let us define the divergence of a vector $\tilde{\sigma} \in V$ so that it verifies $\langle \tilde{\sigma}, \nabla w \rangle_V = -\langle \mathrm{div} \tilde{\sigma}, w \rangle$. By analogy with the continuous setting, we introduce the discrete divergence div_{xy} as the dual operator of ∇_{xy} (which takes into account the Neumann boundary conditions in Ω), i.e., for every $\sigma \in U \times U$ and $w \in U$ we have

$$\langle \mathrm{div}_{xy}\, \sigma, w \rangle = -\langle \sigma, \nabla_{xy} w \rangle_{U \times U}.$$

One can easily check that div_{xy} is given by

$$(\mathrm{div}_{xy}\, \sigma)_{i,j,k} = \begin{cases} \sigma^1_{i,j,k} - \sigma^1_{i-1,j,k} & \text{if } 1 < i < N \\ \sigma^1_{i,j,k} & \text{if } i = 1 \\ -\sigma^1_{i-1,j,k} & \text{if } i = N \end{cases} + \begin{cases} \sigma^2_{i,j,k} - \sigma^2_{i,j-1,k} & \text{if } 1 < j < N \\ \sigma^2_{i,j,k} & \text{if } j = 1 \\ -\sigma^2_{i,j-1,k} & \text{if } j = N \end{cases} \tag{4.4}$$

for every $\sigma = (\sigma^1, \sigma^2) \in U \times U$.

Let us compute the dual operator to $\nabla_s w$, $w \in U$. Let $\sigma^3 \in U$. We have

$$\begin{aligned} \langle \sigma^3, \nabla_s w \rangle &= \sum_{(i,j,k) \in \tilde{\Omega}} \sigma^3_{i,j,k} (\nabla_s w)_{i,j,k} = \sum_{(i,j,k) \in \tilde{\Omega}, 1 \le k < M} \sigma^3_{i,j,k} (w_{i,j,k+1} - w_{i,j,k}) \\ &+ \sum_{(i,j) \in \Omega} \sigma^3_{i,j,0} (w_{i,j,1} - w_{i,j,0}) + \sum_{(i,j) \in \Omega} \sigma^3_{i,j,M} (-w_{i,j,M}) \\ &= \sum_{(i,j,k) \in \tilde{\Omega}, 1 \le k \le M} (\sigma^3_{i,j,k-1} - \sigma^3_{i,j,k}) w_{i,j,k} - \sum_{(i,j) \in \Omega} \sigma^3_{i,j,0} w_{i,j,0}. \end{aligned}$$

Writing $\sigma^3_{i,j,-1} = 0$, we can continue the previous identities as:

$$= \sum_{(i,j,k) \in \tilde{\Omega}, 0 \le k \le M} (\sigma^3_{i,j,k-1} - \sigma^3_{i,j,k}) w_{i,j,k} =: -\langle \nabla_s^- \sigma^3, w \rangle,$$

where

$$(\nabla_s^- \sigma^3)_{i,j,k} \quad = \quad \begin{cases} \sigma^3_{i,j,k} - \sigma^3_{i,j,k-1} & \text{if} \quad 1 \le k \le M \\ \sigma^3_{i,j,0} & \text{if} \quad k = 0 \end{cases} \tag{4.5}$$

We define

$$(\mathrm{div}_{xys}\,(\sigma, \sigma^3))_{i,j,k} = (\mathrm{div}_{xy}\,\sigma)_{i,j,k} + (\nabla_s^- \sigma^3)_{i,j,k}.$$

5. Formulation of discrete nonlinear TV problems in terms of anisotropic total variation . Let

$$\mathcal{U}_M := \{u = (u_{i,j})_{i,j=1}^N \in I\!R^{N\times N} : u_{i,j} \in \{0, \dots, M\}\}.$$

We denote by $|\cdot|_p$ the ℓ_p norm of a vector in $I\!R^2$, $1 \le p \le \infty$.

Let $W : \tilde{\Omega} \to [0, \infty)$. We define

$$\mathcal{R}(u) = \sum_{(i,j)\in\Omega} |(\nabla_{xy} u)_{i,j}|_1 + \sum_{(i,j)\in\Omega} W((i,j), u_{i,j}), \qquad u \in \mathcal{U}_M. \tag{5.1}$$

Let us consider the anisotropy $\phi : \tilde{\Omega} \times I\!R^3 \to [0, +\infty]$ given by

$$\phi((i,j,k), p, p^3) := \begin{cases} |p|_1 + W(i,j,k)|p^3| & \text{if} \quad p^3 \le 0 \\ \\ +\infty & \text{if} \quad p^3 > 0, \end{cases} \tag{5.2}$$

where $(i,j,k) \in \tilde{\Omega}$, $p \in I\!R^2$, $p^3 \in I\!R$. Then if $(i,j,k) \in \tilde{\Omega}$, $\sigma \in I\!R^2$, $\sigma^3 \in I\!R$, by defining

$$\phi^0((i,j,k), \sigma, \sigma^3) = \sup\{\langle (p, p^3), (\sigma, \sigma^3)\rangle : p \in I\!R^2, p^3 \in I\!R : \phi((i,j,k), p, p^3) \le 1\},$$

we have

$$\phi^0((i,j,k), \sigma, \sigma^3) := \begin{cases} |\sigma|_\infty & \text{if} \quad \sigma^3 \ge 0 \\ \\ \max(|\sigma|_\infty, \frac{|\sigma^3|}{W(i,j,k)}) & \text{if} \quad \sigma^3 < 0, \end{cases} \tag{5.3}$$

with the convention that the last maximum is infinity when $W(i,j,k) = 0$. Then for each $(i,j,k) \in \tilde{\Omega}$ we have

$$\begin{aligned} &\{(\sigma, \sigma^3) \in I\!R^2 \times I\!R : \phi^0((i,j,k), \sigma, \sigma^3) \le 1\} \\ &\qquad = \{(\sigma, \sigma^3) \in I\!R^2 \times I\!R : |\sigma|_\infty \le 1,\, (-\sigma^3)^+ \le W(i,j,k)\}. \end{aligned}$$

Let

$$\mathcal{V} := \{\tilde{\sigma} = (\sigma, \sigma^3) \in V : |\sigma_{i,j,k}|_\infty \le 1, (-\sigma^3_{i,j,k})^+ \le W(i,j,k)\ \forall (i,j,k) \in \tilde{\Omega}\},$$

$$\widetilde{\mathcal{A}} := \{w \in U : w_{i,j,k} \in [0,1],\, (\nabla_s w)_{i,j,k} \le 0\ \forall (i,j,k) \in \tilde{\Omega},\, w_{i,j,0} = 1,\ \forall (i,j) \in \Omega\}. \tag{5.4}$$

Let us consider the energy

$$\mathcal{E}(w) = \sum_{(i,j,k)\in\tilde{\Omega}} |(\nabla_{xy} w)_{i,j,k}|_1 + W(i,j,k)|(\nabla_s w)_{i,j,k}| + I_{\{\nabla_s w \le 0\}}(w), \tag{5.5}$$

where $I_{\{\nabla_s w \le 0\}}(w) = 0$ if $\nabla_s w \le 0$, $+\infty$ otherwise. By our definition of the gradient we may write

$$\begin{aligned}\mathcal{E}(w) = & \sum_{(i,j,k)\in\tilde{\Omega}} |(\nabla_{xy} w)_{i,j,k}|_1 + \sum_{(i,j,k)\in\tilde{\Omega}, 1\le k<M} W(i,j,k)|(\nabla_s w)_{i,j,k}| \\ & + \sum_{(i,j)\in\Omega} W(i,j,0)|w_{i,j,1} - 1| + \sum_{(i,j)\in\Omega} W(i,j,M)|w_{i,j,M}| + I_{\{\nabla_s w \le 0\}}(w).\end{aligned} \tag{5.6}$$

Also

$$\begin{aligned}\mathcal{E}(w) &= \sum_{(i,j,k)\in\tilde{\Omega}} |(\nabla_{xy} w)_{i,j,k}|_1 + W(i,j,k)|(\nabla_s w)_{i,j,k}| + I_{\{\nabla_s w \le 0\}}(w) \\ &= \sup_{\{|\sigma_{i,j,k}|_\infty \le 1\}} \sum_{(i,j,k)\in\tilde{\Omega}} \sigma_{i,j,k} \cdot (\nabla_{xy} w)_{i,j,k} \\ &+ \sup_{\{(-\sigma^3_{i,j,k})^+ \le W(i,j,k)\}} \sum_{(i,j,k)\in\tilde{\Omega}} \sigma^3_{i,j,k} \cdot (\nabla_s w)_{i,j,k} \\ &= \sup_{(\sigma,\sigma^3)\in\mathcal{V}} - \sum_{(i,j,k)\in\tilde{\Omega}} (\mathrm{div}_{xy}\,\sigma)_{i,j,k} w_{i,j,k} - \sum_{(i,j,k)\in\tilde{\Omega}} (\nabla_s^- \sigma^3)_{i,j,k} w_{i,j,k} \\ &= \sup_{(\sigma,\sigma^3)\in\mathcal{V}} - \sum_{(i,j,k)\in\tilde{\Omega}} (\mathrm{div}_{xys}\,(\sigma,\sigma_3))_{i,j,k} w_{i,j,k}.\end{aligned}$$

PROPOSITION 5.1. *Let $u \in \mathcal{U}_M$. Let $S(u)$ be the subgraph of u, i.e., $S(u) = \{(i,j,k) \in \tilde{\Omega} : k \le u_{i,j}\}$. We have*

$$\mathcal{R}(u) = \mathcal{E}(\chi_{S(u)}). \tag{5.7}$$

Proof. Let us first observe that

$$\sum_{(i,j,k)\in\tilde{\Omega}} |(\nabla_{xy}\chi_{S(u)})_{i,j,k}|_1 = \sum_{(i,j)\in\Omega} |(\nabla\chi_{S(u)})_{i,j}|_1.$$

To simplify the notation, we use the convention $u(N+1,j) = u(N,j)$ and $(\chi_{S(u)})_{N+1,j,k} =$

$(\chi_{S(u)})_{N,j,k}$, $u(i,N+1) = u(i,N)$, $(\chi_{S(u)})_{i,N+1,k} = (\chi_{S(u)})_{i,N,k}$. We have

$$
\begin{aligned}
&\sum_{(i,j,k)\in\tilde\Omega} |(\nabla_{xy}\chi_{S(u)})_{i,j,k}|_1 = \sup_{\{|\sigma_{i,j,k}|_\infty\le 1\}} \sum_{(i,j,k)\in\tilde\Omega} \sigma_{i,j,k}(\nabla_{xy}\chi_{S(u)})_{i,j,k} \\
&= \sup_{\{|\sigma_{i,j,k}|_\infty\le 1\}} \sum_{(i,j,k)\in\tilde\Omega} (\sigma^1_{i,j,k}[(\chi_{S(u)})_{i+1,j,k} - (\chi_{S(u)})_{i,j,k}] \\
&\qquad\qquad +\sigma^2_{i,j,k}[(\chi_{S(u)})_{i,j+1,k} - (\chi_{S(u)})_{i,j,k}]) \\
&= \sup_{\{|\sigma_{i,j,k}|_\infty\le 1\},\epsilon_{ij},\epsilon^*_{ij}\in\{+1,-1\}} \sum_{(i,j)\in\Omega}\left(\epsilon_{ij}\sum_{m(u)^1_{i,j}<k\le M(u)^1_{i,j}}\sigma^1_{i,j,k} + \epsilon^*_{ij}\sum_{m(u)^2_{i,j}<k\le M(u)^2_{i,j}}\sigma^2_{i,j,k}\right) \\
&= \sum_{(i,j)\in\Omega} |u_{i+1,j}-u_{i,j}| + |u_{i,j+1}-u_{i,j}| = \sum_{(i,j)\in\Omega} |(\nabla_{xy}u)_{i,j}|_1,
\end{aligned}
$$

where $m(u)^1_{i,j} = \min(u_{i,j}, u_{i+1,j})$, $M(u)^1_{i,j} = \max(u_{i,j}, u_{i+1,j})$, $m(u)^2_{i,j} = \min(u_{i,j},\, u_{i,j+1})$, $M(u)^2_{i,j} = \max(u_{i,j}, u_{i,j+1})$.

Writing $(\chi_{S(u)})_{i,j,0} = 1$, $(\chi_{S(u)})_{i,j,M+1} = 0$, we have

$$
\begin{aligned}
\sum_{(i,j,k)\in\tilde\Omega} W(i,j,k)|(\nabla_s\chi_{S(u)})_{i,j,k}| &= \sum_{(i,j,k)\in\tilde\Omega} W(i,j,k)|(\chi_{S(u)})_{i,j,k+1} - (\chi_{S(u)})_{i,j,k}| \\
&= \sum_{(i,j)\in\Omega} W((i,j),u_{i,j}).
\end{aligned}
$$

Both identities prove (5.7). □

Let us prove the coarea formula for $\mathcal{E}$.

THEOREM 5.2. *For any $w \in U$ we have*

$$\mathcal{E}(w) = \int_0^1 \mathcal{E}(\chi_{\{w>t\}})\, dt. \tag{5.8}$$

Proof. Recall that if $a, b \in [0,1]$ we have

$$|a-b| = \int_0^1 |\chi_{[0,a]}(t) - \chi_{[0,b]}(t)|\, dt.$$

The result is immediate if we write $w_{i,j,M+1} = 0$ and use the previous identity. □

PROPOSITION 5.3. *Both problems have a minimum and we have*

$$\min_{w\in\widetilde{\mathcal{A}}} \mathcal{E}(w) = \min_{u\in\mathcal{U}_M} \mathcal{R}(u). \tag{5.9}$$

Proof. Clearly there is a minimum of $\mathcal{R}(u)$ in $\mathcal{U}_M$. Since the anisotropic total variation $\mathcal{E}$ is continuous and $\widetilde{\mathcal{A}}$ is compact in U, we have that there is a minimum of $\mathcal{E}(w)$ in U.

Since any function $u \in \mathcal{U}_M$ determines a function $w = \chi_{S(u)} \in \widetilde{\mathcal{A}}$ such that $\nabla_s w \le 0$, then

$$\min_{u\in\mathcal{U}_M} \mathcal{R}(u) \ge \min_{w\in\widetilde{\mathcal{A}}} \mathcal{E}(w). \tag{5.10}$$

Now, let $w \in \widetilde{\mathcal{A}}$ be a minimum of $\mathcal{E}$. Let $a \in [0,1)$ be such that $\chi_{\{w>a\}}$ is a minimum of $\mathcal{E}$ (this holds for almost any a by the coarea formula). Let $E_k = \{(i,j) \in \Omega : w_{i,j,k} > a\}$, $k \in \{0,\ldots,M\}$. Now, since $E_{k_1} \subseteq E_{k_2}$ whenever $k_1 < k_2$, then we may define

$$u_{i,j} = \sup\{k \in \{0,\ldots,M\} : (i,j) \in E_k\}.$$

Then $\{u \geq k\} = E_k$ for $k \in \{0,\ldots,M\}$ and $u_{i,j} \in \{0,\ldots,M\}$, $\forall (i,j) \in \Omega$. Thus $u \in \mathcal{U}_M$. Since $S(u) = \{w > a\}$, using Proposition 5.1 we have that

$$\mathcal{R}(u) = \mathcal{E}(\chi_{\{w>a\}}) = \min_{w \in \widetilde{\mathcal{A}}} \mathcal{E}(w).$$

We have proved (5.9). □

6. The dual optimization problem for the proximal point method. Let

$$\mathcal{D} := \{\psi \in U : \psi_{i,j,0} = 1 \ \forall (i,j) \in \Omega\}.$$

Let $\psi \in \mathcal{D}$, $\Delta > 0$. Let us consider the problem

$$\min_{w \in \mathcal{D}} \frac{\|w-\psi\|^2}{2\Delta} + \mathcal{E}(w). \tag{6.1}$$

Let us study the optimality condition satisfied by the solution of (6.1). We have

$$\begin{aligned}
&\min_{w \in \mathcal{D}} \frac{\|w-\psi\|^2}{2\Delta} + \mathcal{E}(w) \\
&= \min_{w \in \mathcal{D}} \frac{\|w-\psi\|^2}{2\Delta} + \sum_{(i,j,k)\in\tilde{\Omega}} |(\nabla_{xys} w)_{i,j,k}|_1 + W(i,j,k)|(\nabla_s w)_{i,j,k}| + I_{\{\nabla_s w \leq 0\}}(w) \\
&= \min_{w \in \mathcal{D}} \sup_{\tilde{\sigma} \in \mathcal{V}} \frac{\|w-\psi\|^2}{2\Delta} + \sum_{(i,j,k)\in\tilde{\Omega}} \tilde{\sigma}_{i,j,k} \cdot (\nabla_{xys} w)_{i,j,k} \qquad (*) \\
&= \min_{w \in \mathcal{D}} \sup_{\tilde{\sigma} \in \mathcal{V}} \frac{\|w-\psi\|^2}{2\Delta} - \langle \operatorname{div} \tilde{\sigma}, w\rangle_* - \langle \operatorname{div} \tilde{\sigma}, w\rangle_0 \\
&= \sup_{\tilde{\sigma} \in \mathcal{V}} \min_{w \in \mathcal{D}} \frac{\|w-\psi\|^2}{2\Delta} - \langle \operatorname{div} \tilde{\sigma}, w\rangle_* - \langle \operatorname{div} \tilde{\sigma}, w\rangle_0
\end{aligned}$$

Observe that

$$\langle \operatorname{div} \tilde{\sigma}, w\rangle_0 = \langle \sigma^3, Q\rangle_0,$$

where $Q_{i,j,k} = \delta_{k0}$. Solving

$$\min_{w \in \mathcal{D}} \frac{\|w-\psi\|^2}{2\Delta} - \langle \operatorname{div} \tilde{\sigma}, w\rangle_* - \langle \sigma^3, Q\rangle_0$$

we get

$$w = \psi + \Delta \operatorname{div} \tilde{\sigma} \quad (i,j,k) \in \Omega_*. \tag{6.2}$$

Let us justify the previous computations. For that we recall the following version of the minimax theorem

THEOREM 6.1. *Let $A \subseteq I\!R^m$, $B \subseteq I\!R^k$ be two closed convex sets. Assume that $K : A \times B \to I\!R$ satisfies:*

$$\forall y \in B \quad x \to K(x,y) \text{ is convex and lower semicontinuous,}$$

$$\forall x \in A \quad y \to K(x,y) \text{ is concave and upper semicontinuous.}$$

Assume that for any $y \in B$, $\lambda \in I\!R$, the level sets $\{x \in A : K(x,y) \le \lambda\}$ are compact. Then

$$\inf_A \sup_B K(x,y) = \sup_B \inf_A K(x,y).$$

If we use $A = \mathcal{D} = \{w \in U : w_{i,j,0} = 1 \ \forall (i,j) \in \Omega\}$, $B = \mathcal{V}$,

$$K(w,\tilde{\sigma}) = \frac{\|w-\psi\|^2}{2\Delta} - \langle \operatorname{div} \tilde{\sigma}, w\rangle_* - \langle \sigma^3, Q\rangle_0,$$

then for any $\tilde{\sigma} \in \mathcal{V}$, $\mu \in I\!R$, the sets $\{w \in A : K(w,\tilde{\sigma}) \le \mu\}$ are compact (or empty). Then the previous computations are justified. Notice that this is true even if $\mathcal{V}$ is unbounded.

Introducing w from (6.2) in the above formulas and after some computations we get that

$$\min_{w\in\mathcal{D}} \frac{\|w-\psi\|^2}{2\Delta} + \mathcal{E}(w) = -\frac{\Delta}{2} \min_{\tilde{\sigma}\in\mathcal{V}} \ \|\operatorname{div} \tilde{\sigma}\|_*^2 + \frac{2}{\Delta}\langle \operatorname{div} \tilde{\sigma}, \psi\rangle_* + \frac{2}{\Delta}\langle \sigma^3, Q\rangle_0.$$

PROPOSITION 6.2. *There is a couple $(w,\tilde{\sigma})$ such that $\tilde{\sigma}$ is a solution of the problem*

$$\min_{\tilde{\sigma}\in\mathcal{V}} \ \mathcal{F}(\tilde{\sigma}) \quad \text{where} \quad \mathcal{F}(\tilde{\sigma}) := \|\operatorname{div} \tilde{\sigma}\|_*^2 + \frac{2}{\Delta}\langle \operatorname{div} \tilde{\sigma}, \psi\rangle_* + \frac{2}{\Delta}\langle \sigma^3, Q\rangle_0, \tag{6.3}$$

and $w = \psi + \Delta \operatorname{div} \tilde{\sigma}$, $(i,j,k) \in \Omega_*$, is a solution of

$$\min_{w\in\mathcal{D}} \frac{\|w-\psi\|^2}{2\Delta} + \mathcal{E}(w).$$

Moreover we have

$$\sum_{(i,j,k)\in\tilde{\Omega}} \tilde{\sigma}_{i,j,k} \cdot (\nabla_{xys} w)_{i,j,k} = \sum_{(i,j,k)\in\tilde{\Omega}} |(\nabla_{xys} w)_{i,j,k}|_1 + W(i,j,k)|(\nabla_s w)_{i,j,k}|. \tag{6.4}$$

Observe that as a consequence of (6.4) we have that

$$\sigma^3_{i,j,0}(w_{i,j,1}-1) = \sigma^3_{i,j,0}(\nabla_s w)_{i,j,0} = W(i,j,0)|w_{i,j,1}-1|.$$

Proof. We just have to show that the problem (6.3) has a solution. It suffices to prove that a minimizing sequence is bounded in $\mathcal{V}$. Let $\tilde{\sigma}_n = (\sigma_n, \sigma^3_n)$ be a minimizing sequence of $\mathcal{F}$. Let us write $\sigma^3 = (\sigma^3)^+ - (\sigma^3)^-$, where $(\sigma^3)^+ = \max(\sigma^3, 0)$, $(\sigma^3)^- = \max(-\sigma^3, 0)$. Since $(\sigma^3_n)^-$ and $\mathcal{F}(\tilde{\sigma}_n)$ are bounded, then

$$\|\operatorname{div} \tilde{\sigma}_n\|_*^2 + \frac{2}{\Delta}\langle \operatorname{div} \tilde{\sigma}_n, \psi\rangle_* + \frac{2}{\Delta}\langle (\sigma^3_n)^+, Q\rangle_0$$

is bounded. Hence $\operatorname{div}\tilde{\sigma}_n$ and $\langle(\sigma_n^3)^+, Q\rangle_0$ are bounded. Since σ_n is bounded, we deduce that $\nabla_s\sigma_n^3$ is bounded. Hence, also $\nabla_s(\sigma_n^3)^+$ is bounded. This together with the boundedness of $\langle(\sigma_n^3)^+, Q\rangle_0$, implies that $(\sigma_n^3)^+$ is bounded. Thus, σ_n^3 is bounded and the continuity of $\mathcal{F}$ proves that (6.3) has a solution.

Finally, observe that the identity (6.4) follows from substituting $\tilde{\sigma}$ and w in the identity leading to (*) in the computations at the beginning of this Section. □

7. The narrow band method for the convex relaxation of discrete multi-label problems with TV regularization. Let us use the results proved in Section 3. Let us consider the case where $X = U$ ($m = N \times N \times (M+1)$), $J(w) = \mathcal{E}(w)$ and

$$C = \tilde{\mathcal{A}} = \{w \in U : w_{i,j,k} \in [0,1],\ (\nabla_s w)_{i,j,k} \leq 0\ \forall (i,j,k) \in \tilde{\Omega},\ w_{i,j,0} = 1,\ \forall (i,j) \in \Omega\}.$$

As a consequence of Theorem 5.2 we have:

PROPOSITION 7.1. *The functional $\mathcal{E}(w)$ is a discrete total variation in U.*

The proof of next Proposition has been sketched in Section 2. We give it for the sake of completeness.

PROPOSITION 7.2. *Assume that $\psi \in \mathcal{D}$, $0 \leq \psi \leq 1$. We have $0 \leq w \leq 1$.*

Proof. Let us observe that for any $\alpha \in I\!R$ we have $\mathcal{E}(w \wedge \alpha) \leq \mathcal{E}(w)$ and $\mathcal{E}(w \vee \alpha) \leq \mathcal{E}(w)$. Since $0 \leq \psi \leq 1$, we have that $\|\psi - (w \wedge 1)\| \leq \|\psi - w\|$. Hence

$$\frac{\|\psi - (w \wedge 1)\|}{2\Delta} + \mathcal{E}(w \wedge 1) \leq \frac{\|\psi - w\|}{2\Delta} + \mathcal{E}(w),$$

and by uniqueness of solutions we deduce that $w \wedge 1 = w$. Hence $w \leq 1$. Similarly, using $\|\psi - (w \vee 0)\| \leq \|\psi - w\|$ we obtain that $w \geq 0$. □

Let

$$\mathcal{SG} := \{F \subseteq \tilde{\Omega} : F \text{is the subgraph of } u \in \mathcal{U}_M\}.$$

Notice that if $w = \chi_F$, $F \in \mathcal{SG}$, then $w \in \tilde{\mathcal{A}}$ and $w_{i,j,k} \in \{0,1\}\ \forall (i,j,k) \in \tilde{\Omega}$. Conversely, if $w \in \tilde{\mathcal{A}}$ and $w_{i,j,k} \in \{0,1\}\ \forall (i,j,k) \in \tilde{\Omega}$, then $F := \{(i,j,k) \in \tilde{\Omega} : w_{i,j,k} = 1\} \in \mathcal{SG}$.

Let $L_n, M_n \in \mathcal{SG}$, with $L_n \subset\subset M_n$. We define

$$\tilde{\mathcal{A}}(L_n, M_n) = \{w \in \tilde{\mathcal{A}} : w_{i,j,k} = 1\,\forall (i,j,k) \in L_n,\ w_{i,j,k} = 0\,\forall (i,j,k) \notin M_n\}.$$

Observe that $\tilde{\mathcal{A}} = \tilde{\mathcal{A}}(\{(i,j,0) : (i,j) \in \Omega\}, \tilde{\Omega})$.

LEMMA 7.3. *The sets $\tilde{\mathcal{A}}(L_n, M_n)$ are closed convex sets satisfying property (2.3).*

Proof. If $w \in \tilde{\mathcal{A}}(L_n, M_n)$, we have that $\chi_{\{w \geq t\}} \in \tilde{\mathcal{A}}(L_n, M_n)$ for all $t \in [0,1]$. If $w \in U$, $0 \leq w \leq 1$, is such that $\chi_{\{w \geq t\}} \in \tilde{\mathcal{A}}(L_n, M_n)$ for all $t \in [0,1]$, then $L_n \subseteq \{w \geq t\}$ for all $t < 1$ and $M_n^c \subseteq \{w < t\}$ for all $t > 0$. Thus $w_{i,j,k} = 1$ for all $(i,j,k) \in L_n$, $w_{i,j,k} = 0$ for all $(i,j,k) \notin M_n$. Since

$$w(i,j,k) = \int_0^1 (\chi_{\{w \geq t\}})_{i,j,k}\, dt$$

it follows that $\nabla_s w \leq 0$. Hence $w \in \tilde{\mathcal{A}}(L_n, M_n)$. □

Then Theorem 3.2 proves that the narrow band method converge to a local minimum of the problem:

$$\min_{w \in \tilde{\mathcal{A}}} \mathcal{E}(w). \tag{7.1}$$

In order to have a narrow band method we adapt the data domain at each Step. Start with $w^0 = \overline{w}^0 \in \tilde{\mathcal{A}}$ with $\mathcal{E}(w^0) < \infty$, $\tilde{\mathcal{A}}(L_0, M_0) = \tilde{\mathcal{A}}$. Then at each Step we assume that we defined $L_n, M_n \in \mathcal{SG}$, with $L_n \subset\subset M_n$, that we have $w^n \in \tilde{\mathcal{A}}(L_n, M_n)$ and we solve the problem

$$\min_{w \in \tilde{\mathcal{A}}(L_n, M_n)} \frac{1}{2\Delta} \|w - w^n\|^2 + \mathcal{E}(w). \tag{7.2}$$

Let $\overline{w}^{n+1}$ be the unique solution of (7.2). Let us denote $T_t(w) = \chi_{\{w>t\}}$, $w \in \tilde{\mathcal{A}}$.

Given $\overline{w}^{n+1}$ we may find t_{n+1} such that $\mathcal{E}(T_{t_{n+1}} \overline{w}^{n+1}) \leq \mathcal{E}(\overline{w}^{n+1}) + \epsilon_{n+1}$ where $\sum_{n=1}^{\infty} \epsilon_n < \infty$. The choice of $\mathcal{E}(T_{t_{n+1}} \overline{w}^{n+1})$ is described with further detail in Lemmas 3.3 and 3.4.

We define

$$w^{n+1} = T_{t_{n+1}}(\overline{w}^{n+1}).$$

We define $L_{n+1}, M_{n+1} \in \mathcal{SG}$ such that $L_{n+1} \subset\subset \{w^{n+1} = 1\} \subset\subset M_{n+1}$. We can use a construction which at each step gives us unique such sets L_n, M_n. Now the sequence of problems (7.2) is completely determined.

THEOREM 7.4. *For n large enough the solution of (7.2) becomes a constant and binary function. Thus the algorithm converges in finitely many steps to a local minimum of $\mathcal{E}$.*

We also have

LEMMA 7.5. *Assume that $\epsilon \leq \frac{1}{2}$ and let n be such that (3.6) holds. Then the infimum of $\{\mathcal{E}(\chi_{\{\overline{w}^{n+1} > t\}}) : t \in [\min \overline{w}^{n+1}, \max \overline{w}^{n+1})\}$ is attained at one of the values $\mathcal{E}(\chi_{\{\overline{w}^{n+1} > \min \overline{w}^{n+1}\}})$ or $\mathcal{E}(\chi_{\{\overline{w}^{n+1} \geq \max \overline{w}^{n+1}\}})$.*

7.1. Formulation of the energy and constraint in terms of an anisotropic total variation. Let $\lambda > 0$. Let us consider the anisotropy

$$\phi_\lambda((i,j,k), p, p^3) := |p|_1 + W(i,j,k)|p^3| + \lambda (p^3)^+ \tag{7.3}$$

Then

$$\phi^0_\lambda((i,j,k), \sigma, \sigma^3) := \begin{cases} \max(|\sigma|_\infty, \dfrac{|\sigma^3|}{W(i,j,k) + \lambda}) & \text{if} \quad \sigma^3 \geq 0 \\ \\ \max(|\sigma|_\infty, \frac{|\sigma^3|}{W(i,j,k)}) & \text{if} \quad \sigma^3 < 0. \end{cases} \tag{7.4}$$

with the convention that the last maximum is infinity when $W(i,j,k) = 0$. Then for each $(i,j,k) \in \tilde{\Omega}$ we have

$$\mathcal{V}_\lambda := \{(\sigma, \sigma^{N+1}) \in V : \phi^0_\lambda((i,j,k), \sigma, \sigma^3) \leq 1\}$$

$$= \{(\sigma, \sigma^3) \in V : |\sigma_{i,j,k}|_\infty \leq 1, \, |\sigma^3_{i,j,k}| \leq W(i,j,k) \text{ if } \sigma^3_{i,j,k} < 0,$$

$$|\sigma^3_{i,j,k}| \leq W(i,j,k) + \lambda \text{ if } \sigma^3_{i,j,k} \geq 0\}.$$

Let us consider the energy

$$\mathcal{E}_\lambda(w) = \sum_{(i,j,k) \in \tilde{\Omega}} |(\nabla_{xy} w)_{i,j,k}|_1 + W(i,j,k)|(\nabla_s w)_{i,j,k}| + \lambda((\nabla_s w)_{i,j,k})^+. \tag{7.5}$$

By our definition of the gradient we may write

$$
\begin{aligned}
\mathcal{E}_\lambda(w) = & \sum_{(i,j,k)\in\tilde{\Omega}} |(\nabla_{xy} w)_{i,j,k}|_1 + \sum_{(i,j,k)\in\tilde{\Omega}, 1\le k<M} W(i,j,k)|(\nabla_s w)_{i,j,k}| + \lambda((\nabla_s w)_{i,j,k})^+ \\
& + \sum_{(i,j)\in\Omega} W(i,j,0)|w_{i,j,1}-1| + \lambda(w_{i,j,1}-1)^+ \\
& + \sum_{(i,j)\in\Omega} W(i,j,M)|w_{i,j,M}| + \lambda(-w_{i,j,M})^+.
\end{aligned}
\tag{7.6}
$$

LEMMA 7.6. *We have that* $\kappa := \max_{0\le w\le 1} \|\partial\mathcal{E}_\lambda(w)\|_\infty < \infty$.

Indeed $\kappa \le \|\nabla_{xy}\| + \|\nabla_s\|(\|W\|_\infty + \lambda)$.

Let us consider

$$\tilde{\mathcal{A}}_\lambda = \{w \in U : 0 \le w_{i,j,k} \le 1 \;\forall (i,j,k)\in\tilde{\Omega},\; w_{i,j,0} = 1\,\forall (i,j)\in\Omega\}.$$

Then we apply the narrow band strategy to minimize

$$\min_{w\in\tilde{\mathcal{A}}_\lambda} \mathcal{E}_\lambda(w). \tag{7.7}$$

Again, in order to have a narrow band method we adapt the data domain at each Step. Start with $w^0 = \overline{w}^0 \in \tilde{\mathcal{A}}_\lambda$ with $\mathcal{E}(w^0) < \infty$, $\tilde{\mathcal{A}}_\lambda(L_0, M_0) = \tilde{\mathcal{A}}_\lambda$. Then at each Step we have defined $L_n, M_n \in \mathcal{SG}$, with $L_n \subset\subset M_n$, we have $w^n \in \tilde{\mathcal{A}}_\lambda(L_n, M_n)$, and we solve the problem

$$\min_{w\in\tilde{\mathcal{A}}_\lambda(L_n,M_n)} \frac{1}{2\Delta}\|w-w^n\|^2 + \mathcal{E}_\lambda(w). \tag{7.8}$$

Let $\overline{w}^{n+1}$ be the unique solution of (7.8). Let us denote $T_t(w) = \chi_{\{w>t\}}$, $w\in\tilde{\mathcal{A}}_\lambda$.

Given $\overline{w}^{n+1}$ we may find t_{n+1} such that $\mathcal{E}(T_{t_{n+1}}\overline{w}^{n+1}) \le \mathcal{E}(\overline{w}^{n+1}) + \epsilon_{n+1}$ where $\sum_{n=1}^\infty \epsilon_n < \infty$.

Notice that we may replace $\{\overline{w}^{n+1} > t_{n+1}\}$ by its connected component A_{n+1} which contains $\{(i,j,0) : (i,j)\in\Omega\}$. Indeed $\mathcal{E}_\lambda(\chi_{A_{n+1}}) \le \mathcal{E}_\lambda(\chi_{\{\overline{w}^n > t_n\}})$. We define

$$w^{n+1} = \chi_{A_{n+1}}.$$

We define $L_{n+1}, M_{n+1} \in \mathcal{SG}$ such that $L_{n+1} \subset\subset A_{n+1} \subset\subset M_{n+1}$. Now the sequence of problems (7.8) is completely determined.

LEMMA 7.7. *Let* $\overline{w}_\lambda$ *be a minimum of (7.7). If* $\lambda > 4M$*, then* $\nabla_s \overline{w}_\lambda \le 0$.

Observe that if we take $\lambda_0 > 4M$, $\kappa = \kappa(\lambda_0) := \|\nabla_{xy}\| + \|\nabla_s\|(\|W\|_\infty + \lambda_0)$ and $\Delta = \frac{1}{\kappa(\lambda_0)}$, then Lemmas 3.4 and 7.7 hold. Then we converge to a local minimum of $\mathcal{E}$ which is the graph of a solution of $\mathcal{R}$.

Proof. Let $\overline{w}_\lambda$ be a minimum of (7.7). Then its level sets are also minima of (7.7) and are binary functions. If we prove that $\nabla_s \chi_{\{\overline{w}_\lambda > t\}} \le 0$, then it follows that $\nabla_s \overline{w}_\lambda \le 0$. Thus we may assume that $\overline{w}_\lambda$ is a binary function, i.e., $\overline{w}_\lambda = \chi_F$ where $F \in \mathcal{SG}$. If $\nabla_s \overline{w}_\lambda \le 0$ is not true, then there exists $(i_0, j_0) \in \Omega$ and an interval $I := \{k_1+1, \ldots, k_2-1\}$ of $\{0, \ldots, M\}$ with $k_1 + 1 < k_2$ such that $(i_0, j_0, k_1) \in F$,

$(i_0, j_0, k_2) \in F$ and $(i_0, j_0, k) \notin F$ for any $k \in I$. Then we define $F_I = F \cup \{(i_0, j_0, k) : k \in I\}$ Then

$$\sum_{(i,j,k)\in\tilde{\Omega}} W(i,j,k)|(\nabla_s \chi_{F_I})_{i,j,k}| + \lambda((\nabla_s \chi_{F_I})_{i,j,k})^+$$

$$\leq \sum_{(i,j,k)\in\tilde{\Omega}} W(i,j,k)|(\nabla_s \chi_F)_{i,j,k}| + \lambda((\nabla_s \chi_F)_{i,j,k})^+ - \lambda.$$

On the other hand

$$\sum_{(i,j,k)\in\tilde{\Omega}} |(\nabla_{xy}\chi_{F_I})_{i,j,k}|_1 \leq \sum_{(i,j,k)\in\tilde{\Omega}} |(\nabla_{xy}\chi_F)_{i,j,k}|_1 + 4M.$$

Hence

$$\mathcal{E}_\lambda(\chi_{F_I}) - \mathcal{E}_\lambda(\chi_F) \leq 4M - \lambda < 0.$$

This contradiction proves that $\nabla_s \overline{w}_\lambda \leq 0$. ∎

8. Experiments. We applied the narrow band algorithm to two different multi-label problems: disparity estimation and segmentation.

8.1. Disparity estimation. This first problem consists in finding the disparity between two rectified images, as the ones shown in Figure 8.1. Denoting I_1 and I_2 the two images with domain $\Omega = \{1, \ldots, N\}^2$, we aim at seeking the disparity u of image I_1 with respect to image I_2, which takes its values into the discrete set $\{l_0, \cdots, l_n\}$. Let $\mathcal{U} := \{u = (u_{i,j})_{i,j=1}^N : u_{i,j} \in \{l_0, \cdots, l_n\}\}$. The problem can be formulated as:

$$\min_{u\in\mathcal{U}} \sum_{(i,j)\in\Omega} |(\nabla_{xy}u)_{i,j}|_1 + \alpha \sum_{(i,j)\in\Omega} W(i,j,u_{i,j}), \tag{8.1}$$

where $\alpha > 0$, and the cost W for a pixel $(i,j) \in \Omega$ and a disparity $u_{i,j}$ is defined as:

$$W(i,j,u_{i,j}) = (I_1(i,j) - I_2(i + u_{i,j}, j))^2.$$

We applied this process to the disparity estimation of the images shown in Figure 8.1 with 16 labels and $\alpha = 40$.

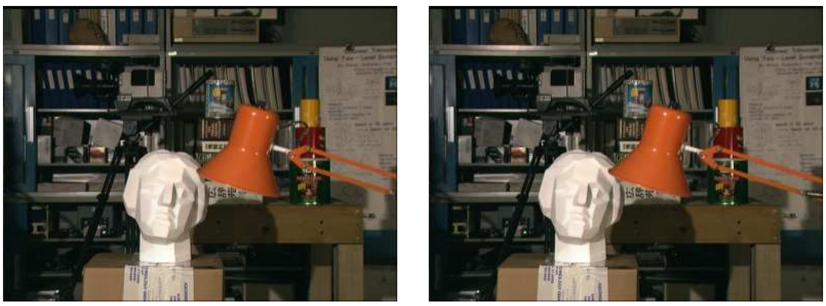


Fig. 8.1. *A pair of rectified images.*

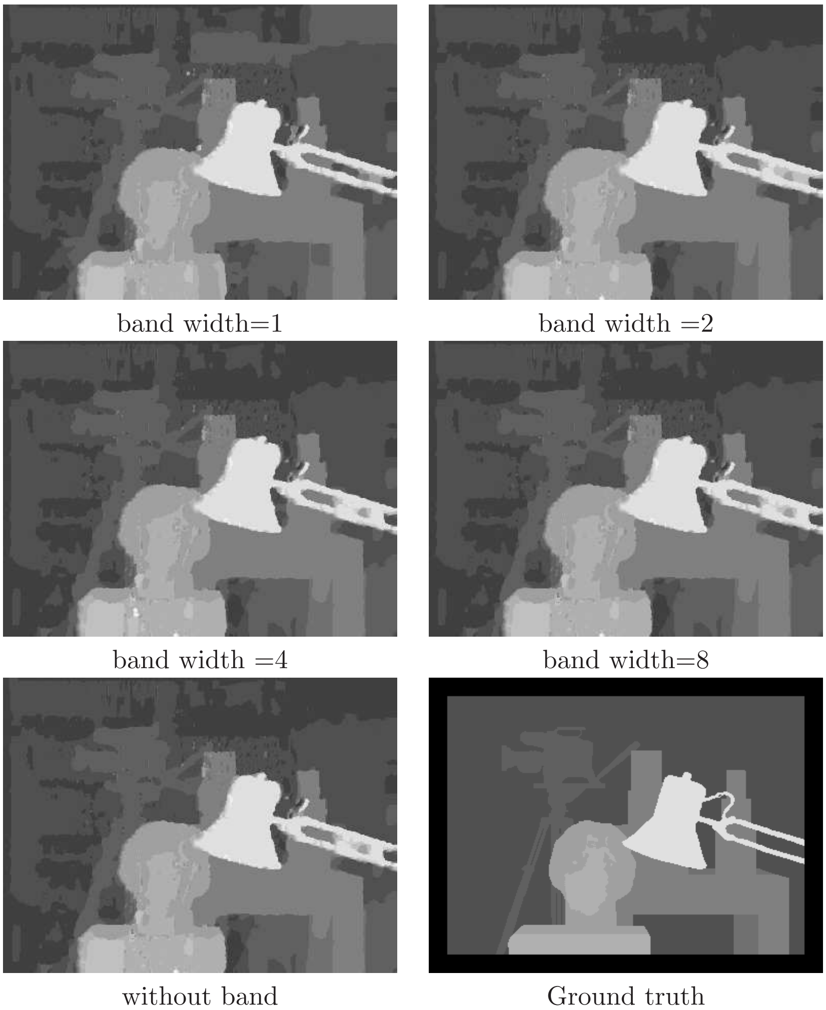


FIG. 8.2. *Disparity estimation for increasing value of band size.*

In Figure 8.2, we show the disparities obtained by the process for different sizes of the band. The result without using a narrow band is also given. We can see that for increasing band size, the result gets closer to the one obtained without band.

Table 8.3 gives a quantitative comparison of the obtained results with respect to the ground truth using a threshold of 1 for wrong pixels. We can see from these results that using a narrow band does not deteriorate the estimation, even for small band widths. It even sometimes gives a better solution while not reaching the global minimum of the energy. Indeed, as the disparity problem is not perfectly described by the expression (8.1), reaching the global minimum does not guarantee a better solution.

Let us now discuss these results in terms of energy and speed. In Figure 8.4, the values of the minimum of the energy of problem (8.1) reached for the different band sizes are drawn. We observe that the values of the energy obtained for any band width are close to the energy obtained without band (the differences of the energies

Method	Error (in %)
band width =1	2.25
band width=2	2.43
band width=4	2.50
band width=8	2.32
without band	2.33

FIG. 8.3. *Percentage of wrong pixels with a threshold of 1 for the different estimations.*

obtained with respect to the result without band, in relative terms, are smaller than a 1% in all cases), and that increasing the band width allows recovering the global minimum reached by the method without band.

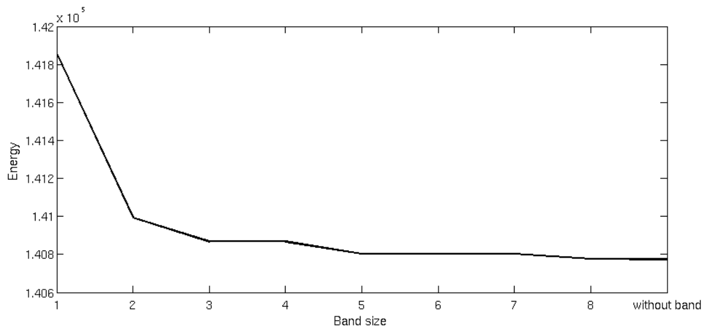


FIG. 8.4. *Value of the energy (8.1) reached for different sizes of the band.*

Concerning the speed analysis, we computed first the global minimum of the problem (8.1) thanks to the method without band and evaluated its energy value E_{min}. Then we applied the process for different sizes of the band (and with no band) until reaching an energy E sufficiently close to the global minimum: $||E - E_{min}|| \leq 0.01E_{min}$. In figure 8.5 we show the time necessary to reach this energy threshold. As it can be observed, the computational time increases with the width of the band until it reaches the time corresponding to the use of the whole domain. Notice also that in the case of the narrow band, some additional computations are required in order to compute the best threshold cut.

Let us finally underline that in our experiments, the unknown u was initialized in a particular way. For each pixel (i, j) we initially choose $u_{i,j}$ so that $W((i, j), u_{i,j})$ is a minimum. This is equivalent to minimize the functional (8.1) without considering the regularization term. Without this "good" initialization, the narrow band process can be stuck into bad local minima, as illustrated in Figure 8.6.

8.2. Segmentation. In the case of segmentation, as in [14, 8], we solve the Mumford-Shah functional for piecewise constant functions. Given an image I whose domain is Ω, we look for the optimal image partition Ω_k, $k = 0, \ldots, M$, where a cost $W_2((i, j), k)$ is defined for each pixel $(i, j) \in \Omega$ and each label $k = 0, \ldots, M$. Associating a different gray level (or color) c_k to each label, such a cost can be defined as:

$$W_2(i, j, k) = (I(i, j) - c_k)^2.$$

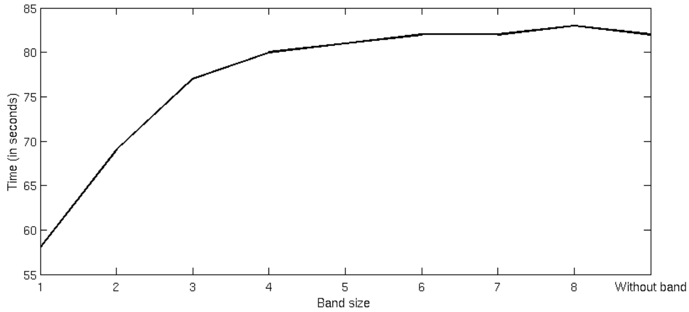


FIG. 8.5. *Time necessary to reach an energy value close to the global minimum of the energy (8.1) for different sizes of the band.*

The Mumford-Shah problem can then be formulated as the minimization of the following functional for a label function $u : \Omega \mapsto \{0, \dots, M\}$.

$$\min_u \sum_{(i,j)\in\Omega} \delta(|(\nabla_{xy} u)_{i,j}|) + \alpha \sum_{(i,j)\in\Omega} W_2(i,j,u_{i,j}), \tag{8.2}$$

where $\alpha > 0$ and $\delta(|(\nabla_{xy} u)_{i,j}|) = 1$ if $(\nabla_{xy} u)_{i,j} \neq 0$, 0 otherwise. This is the Potts model regularization. The resulting partition is given by $\Omega_k := \{(i,j) : u_{i,j} = k\}$, $k = 0, \dots, M$. The convex relaxation of this model has been proposed and studied in [8, 14]. Let us recall that, in this case, the relaxed convex problem for the embedding function w defined as $w(i,j,k) = \chi_{\{u(i,j)\geq k\}}$ can be formulated as [14]:

$$\min_w \left(\max_{\sigma\in K} \sum_{(i,j,k)\in\tilde\Omega} \sigma_{i,j,k} (\nabla_{xy} w)_{i,j,k} + \max_{-\sigma^3_{i,j,k} \leq W_2(i,j,k)} \sum_{(i,j,k)\in\tilde\Omega, 1\leq k<M} (\nabla_s w)_{i,j,k} \sigma^3_{i,j,k} \right),$$

where σ now belongs to the set K defined as:

$$K = \{\sigma = (\sigma_{i,j,k})_{(i,j,k)\in\tilde\Omega} : \sigma_{i,j,k} \in I\!R^2 \ \forall (i,j,k) \in \tilde\Omega,$$
$$\left| \sum_{k_1 \leq k \leq k_2} \sigma_{i,j,k} \right|_1 \leq 1, \ \forall (i,j) \in \Omega \text{ and } \forall (k_1,k_2) \text{ with } 0 \leq k_1 \leq k_2 \leq M\}.$$

The projection of σ into this set is computed as in [14] using Dykstra's algorithm.

The methods developed in this paper can be applied to this relaxed convex model. We applied this process to segment the images shown in Figure 8.7. For a grey level image I, we used $\alpha = 10$ and 16 labels whose corresponding reference labels c_i are uniformly distributed between the minimum and the maximum gray levels.

In Figure 8.8, we show the segmentation of the first image in Figure 8.7 obtained with the narrow band algorithm for different sizes of the band. The result obtained without using a narrow band is also given. The segmentations obtained are quite similar. It can nevertheless be noticed that some small structures disappear when the band size is increased. Indeed, when the band size is too small, the process is too local.

As for the disparity estimation problem, we checked numerically in Figure 8.9 that the energy values reached by the narrow band method converge to the global energy

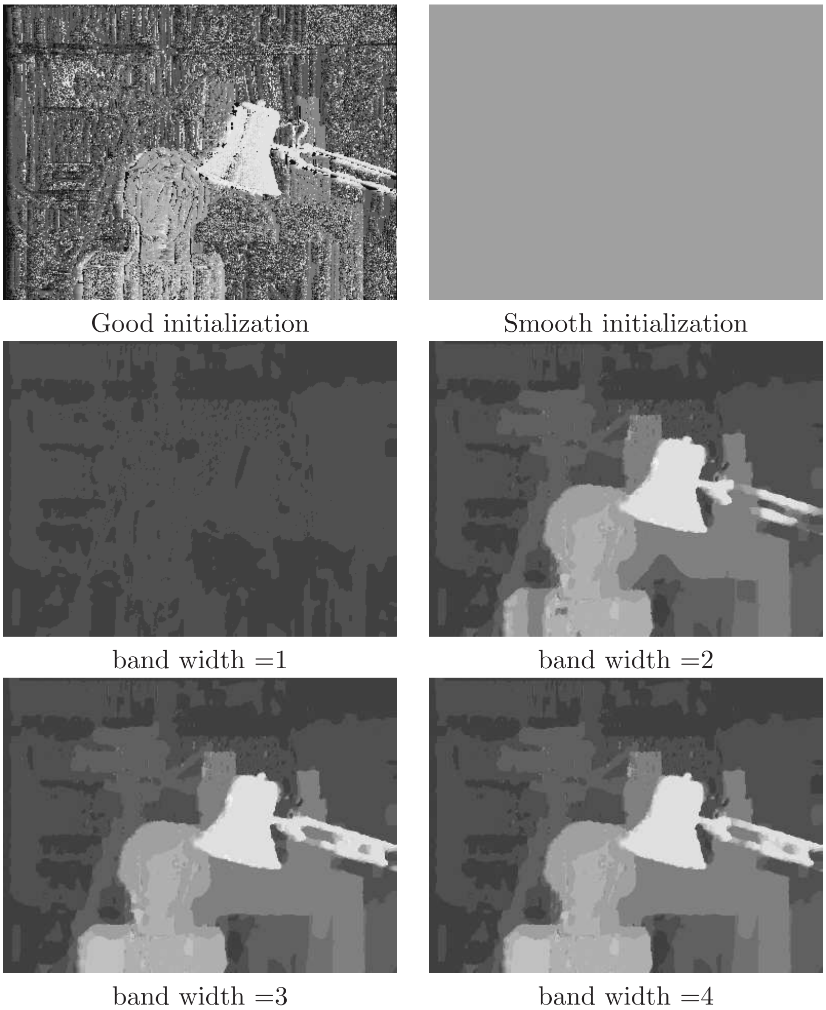


FIG. 8.6. *Solution obtained with the narrow band method with a smooth initialization.*

minimum of problem (8.2). The computation time necessary to reach a neighborhood of this global minimum is also illustrated in Figure 8.10 and still increases with band size.

We finally show in Figures 8.11 and 8.12, the segmentations obtained on the two last images of Figure 8.7.

Acknowledgements. This work was partially funded by Mediapro through the Spanish project CENIT-2007-1012 i3media and by the Centro para el Desarrollo Tecnológico Industrial (CDTI), within the Ingenio 2010 initiative. V. Caselles and P. Gargallo acknowledge partial support by IP project "2020 3D Media: Spacial Sound and Vision", Financed by EC. P. Gargallo and N. Papadakis acknowledge support from the Torres Quevedo program of the Ministerio de Educación y Ciencia in Spain. V. Caselles also acknowledges partial support by MICINN project, reference MTM2009-08171, by GRC reference 2009 SGR 773 and by "ICREA Acadèmia" prize for excellence in research funded both by the Generalitat de Catalunya.

Fig. 8.7. *Images used for segmentation.*

REFERENCES

[1] G. Alberti, G. Bouchitté, and D. Dal Maso, *The calibration method for the Mumford-Shah functional and free-discontinuity problems*, Calc. Var. Partial Differential Equations, 16(2003), pp. 299–333.

[2] F. Alter, V. Caselles, and A. Chambolle, *A characterization of convex calibrable sets in* $\mathbb{R}^N$, Math. Ann., 332(2005), pp. 329–366.

[3] M. Amar and G. Bellettini, *A notion of total variation depending on a metric with discontinuous coefficients*, Ann. Inst. Henri Poincaré, 11(1994), pp. 91–133.

[4] L. Ambrosio, N. Fusco and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs, 2000.

[5] Y. Boykov, O. Veksler, and R. Zabih, *Fast approximate energy minimization via graph cuts*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2001), pp. 1222–1239.

[6] X. Bresson, S. Esedoglu, P. Vandergheynst, J.P. Thiran, and S. Osher, *Fast global minimization of the active contour/snake model*, Journal of Mathematical Imaging and Vision, 28(2007), pp. 151–167.

[7] H. Brezis, *Operateurs Maximaux Monotones*, North Holland, Amsterdam, 1973.

[8] E.S. Brown, T.F. Chan, and X. Bresson, *Convex formulation and exact global solutions for multi-phase piecewise constant Mumford-Shah image segmentation*, UCLA CAM Report 09–66, 2009.

[9] A. Almansa, C. Ballester, V. Caselles, and G. Haro. *A TV based restoration model with local constraints*, Journal Scientific Computing, 34(2008), pp. 209–236.

[10] V. Caselles, G. Facciolo, and E. Meinhardt, *Anisotropic Cheeger sets and applications*, SIAM Journal on Imaging Sciences, 2(2009), pp. 1211–1254.

[11] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20(2004), pp. 89–97.

[12] A. Chambolle, *Total variation minimization and a class of binary MRF models*, In Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science, 2005, pp. 136–152.

[13] A. Chambolle and J. Darbon, *On Total variation minimization and Surface Evolution using Parametric Maximum Flows*, Preprint, 2009.

[14] A. Chambolle, D. Cremers, and T.Pock, *A convex approach for computing minimal partitions*, Preprint R.I. 649, CMA Ecole Polytechnique, 2008.

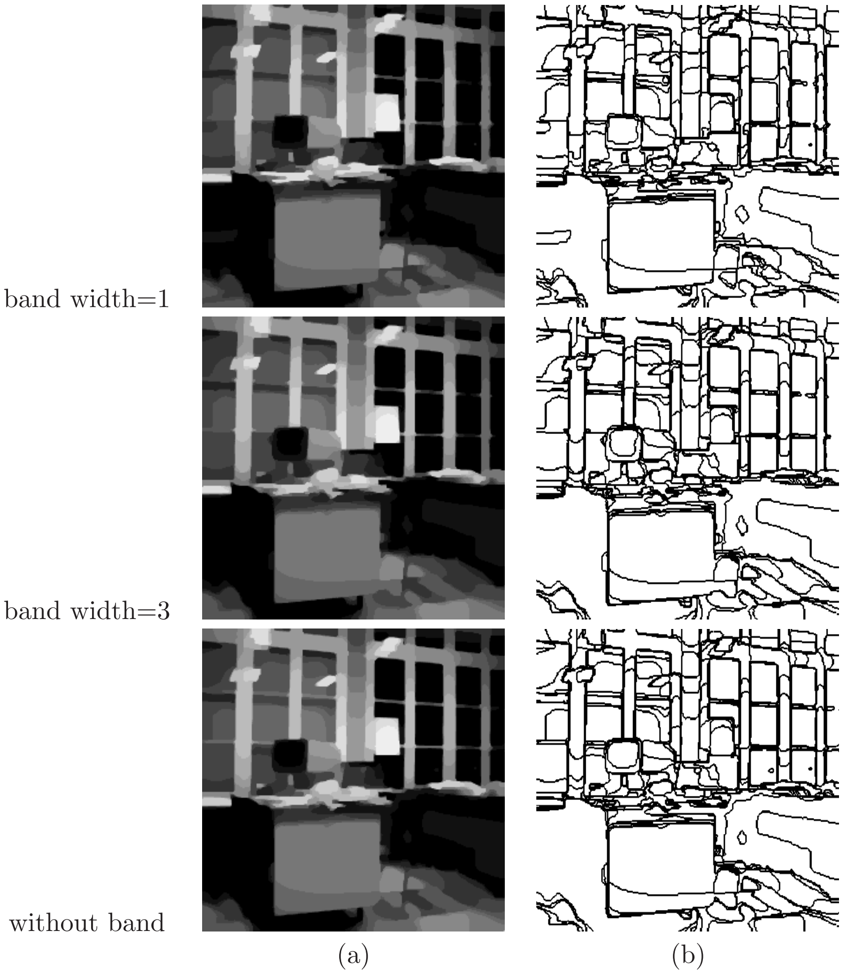


Fig. 8.8. *(a) Segmentations obtained for different sizes of the band. (b) Boundaries of segmented regions.*

[15] T.F. Chan, S. Esedoglu, and M. Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM Journal on Applied Mathematics 66(2006), pp. 1632–1648.

[16] D.M. Greig, B.T. Porteous, and A.H. Seheult, *Exact maximum a posteriori estimation for binary images*, Journal of the Royal Statistical Society. Series B (Methodological) 51(1989), pp. 271–279.

[17] H. Ishikawa, *Exact optimization for Markov random fields with convex priors*, IEEE Transactions on Pattern Analysis and Machine Intelligence 25(2003), pp. 1333–1336.

[18] V. Kolmogorov and R. Zabih, *What energy functions can be minimized via graph cuts?*, IEEE Trans. Pattern Analysis and Machine Intelligence, 2(2004), pp. 147–159.

[19] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr, *Convex multi-class image labeling by simplex-constrained total variation*, Technical Report, IPA, HCI, Dept. of Mathematics and Computer Science, Univ. Heidelberg, October 2008.

[20] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers, *A convex formulation of continuous multi-label problems*, In European Conference on Computer Vision (ECCV), Marseille, France, October 2008.

[21] T. Pock, M. Unger, D. Cremers, and H. Bischof, *Fast and Exact Solution of Total Variation Models on the GPU*, CVPR Workshop on Visual Computer Vision on GPU's June 2008.

[22] T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control and

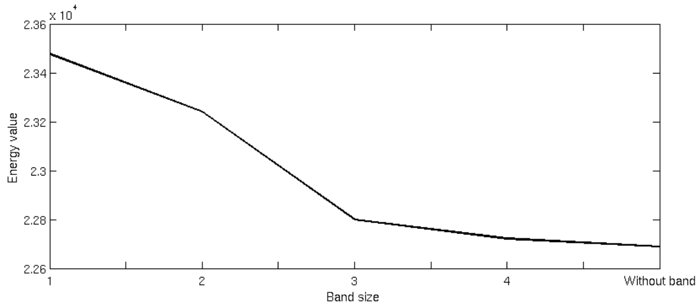


FIG. 8.9. *Value of the energy (8.2) reached for different sizes of the band.*

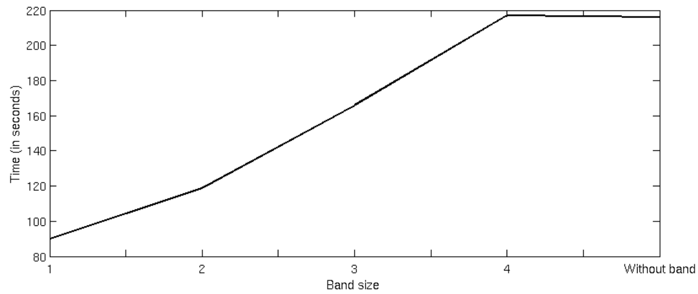


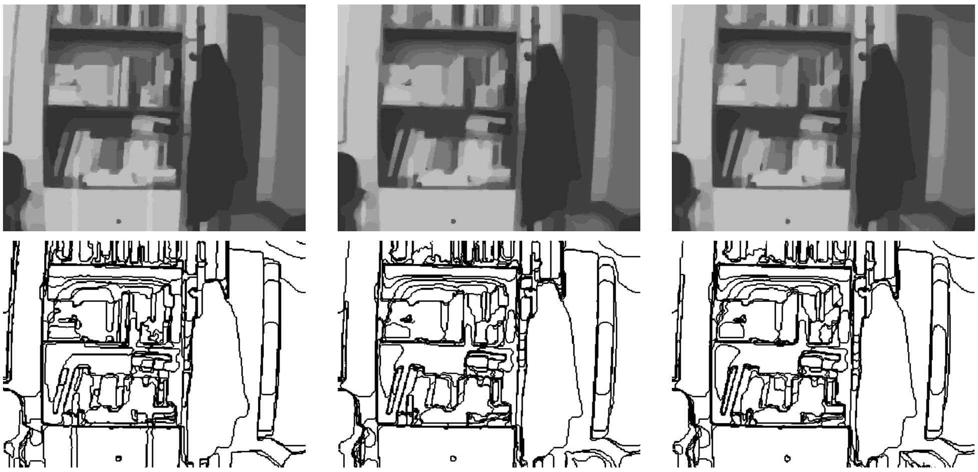
FIG. 8.10. *Time required to reach an energy value close to the global minimum of the energy (8.2) for different sizes of the band.*

Optimization, 14(1976), pp. 877–898.

[23] O. Veksler, *Efficient graph-based energy minimization methods in computer vision*, PhD Thesis, Cornell University, Ithaca, NY, July 1999.

[24] C. Zach, D. Gallup, J.M. Frahm, and M. Niethammer, *Fast global labeling for real-time stereo using multiple plane sweeps*, Proceedings of Vision, Modeling, and Visualization 2008, October 8-10, 2008, Konstanz, Germany, 243, 2008.

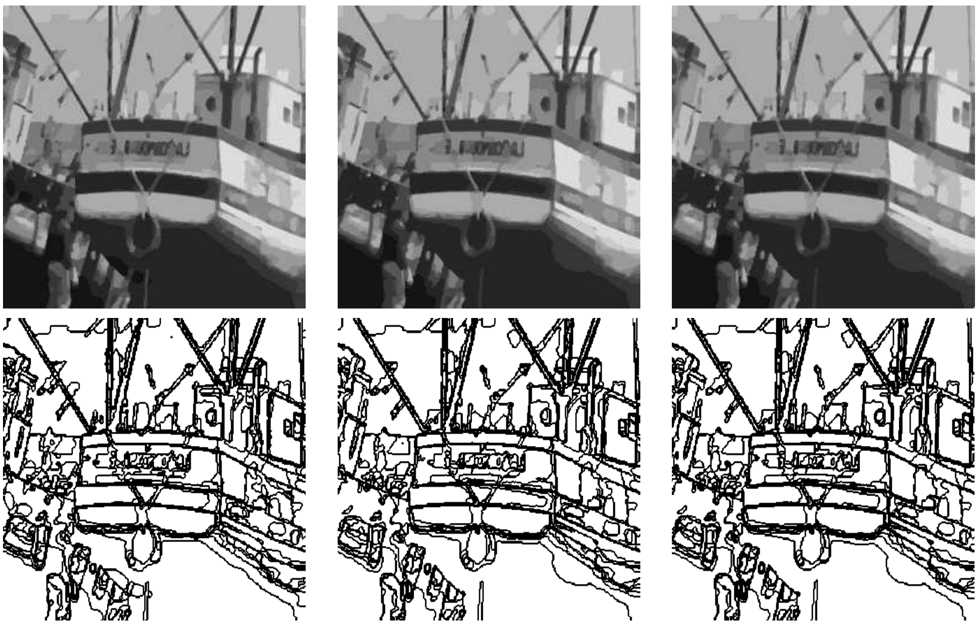
[25] M. Zhu and T.F. Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, Preprint 2008.



(a) (b)

band width =1 band width =3 without band

FIG. 8.11. *(a) Segmentations obtained for different sizes of the band. (b) Boundaries of segmented regions.*



(a) (b)

band width =1 band width =3 without band

FIG. 8.12. *(a) Segmentations obtained for different sizes of the band. (b) Boundaries of segmented regions.*