TELECOM PARIS
école nationale supérieure des télécommunications

A fast and exact algorithm for total variation minimization

Un algorithme rapide et exact de la minimisation de la variation totale

Jérôme Darbon
Marc Sigelle

2005D002

janvier 2005

Département Traitement du Signal et des Images
Groupe Traitements Statistiques et Applications aux Communications

Dépôt légal : 2005 – 1er trimestre
Imprimé à l'Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 0751-1345 ENST D (Paris) (France 1983-9999)

Ecole Nationale Supérieure des Télécommunications
Groupe des Ecoles des Télécommunications - membre de ParisTech
46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.enst.fr
Département TSI

A Fast and Exact Algorithm for Total Variation Minimization

Jérôme Darbon[1,2] Marc Sigelle[2]

January 20^{th}, 2005

[1] EPITA Research and Development Laboratory (LRDE),
14-16 rue Voltaire, F-94276 Le Kremlin-Bicêtre, France,
jerome.darbon@{lrde.epita.fr, enst.fr}

[2] ENST / LTCI CNRS UMR 5141,
46 rue Barrault, F-75013 Paris, France
marc.sigelle@enst.fr

A paper based on this technical report has been accepted to IbPria'2005 and published by Springer LNCS Series

Abstract

This paper deals with the minimization of the total variation under a convex data fidelity term. We propose an algorithm which computes an exact minimizer of this problem. The method relies on the decomposition of an image into its level sets. Using these level sets, we map the problem into optimizations of independent binary Markov Random Fields. Binary solutions are found thanks to graph-cut techniques and we show how to derive a fast algorithm. We also study the special case when the fidelity term is the L^1-norm. Finally we provide some experiments.

Un algorithme rapide et exact de minimisation de la variation totale

Jérôme Darbon[1,2] Marc Sigelle[2]

20 janvier 2005

[1] Laboratoire de recherche et développement de l'EPITA (LRDE),
14-16 rue Voltaire, F-94276 Le Kremlin-Bicêtre, France,
jerome.darbon@{lrde.epita.fr, enst.fr}

[2] ENST / LTCI CNRS UMR 5141,
46 rue Barrault, F-75013 Paris, France
marc.sigelle@enst.fr

Un papier utilisant les résultats de ce rapport a été accepté au congrès IbPria'2005 et publié chez Springer LNCS

Résumé

Ce papier traite de la minimisation de la variation totale avec une attache aux données convexe. Nous proposons un algorithme qui calcule un minimiseur exact de ce problème. Notre méthode repose sur la décomposition d'une image en ses ensembles de niveaux. En utilisant ces ensembles de niveaux, nous reformulons le problème en termes de champs de Markov binaire indépendants. Les solutions binaires de ces problèmes sont calcules grâce un algorithme de coupure minimale et nous en dduisons un algorithme rapide pour résoudre le problème original. Nous étudions également le cas particulier d'une attache aux données modélisée par la norme L^1. Enfin, nous présentons quelques résultats.

1 Introduction

Minimization of the total variation (tv) for image reconstruction is of great importance for image processing applications [1, 17, 19, 21, 22]. It has been shown that these minimizers live in the space of bounded variation [9] which preserves edges and allows for sharp boundaries. In this paper we propose a new and fast algorithm which computes an exact solution of tv minimization-based problems.

Assume u is an image defined on Ω then its total variation is $tv(u) = \int_\Omega |\nabla u|$, where the gradient is taken in the distributional sense. A classical approach to minimize tv is achieved by a gradient descent [24] which yields the following evolution equation $\frac{\partial u}{\partial t} = \text{div}\left(\frac{\nabla u}{|\nabla u|+\epsilon}\right)$. To avoid division by zero, ϵ is set to a small positive value. In [5], Chambolle reformulates tv minimization problem using duality. Using this formulation he proposes a fast algorithm. In [19], Pollak *et al.* present a fast algorithm which provide the exact solution in one dimension. However only an approximation is available in higher dimensions. After a discretization, tv minimization can be reformulated as a minimization problem involving a Markov Random Field (MRF). In [4], Boykov *et al.* present a fast approximation minimization algorithm based on graph cuts for MRF. An algorithm which computes an exact solution for MRF where the prior is convex is presented in [12]. It is also based on graph-cuts.

In this paper, we assume u and v are two images defined on Ω. Thus we are interested in minimizing the following functional:

$$E(u) = \int_\Omega f\left(u(x),\ v(x)\right) dx + \beta \int_\Omega |\nabla u| \,. \tag{1}$$

We assume that the attachment to data term is a convex function of $u(.)$, such as: $f\left(u(x),\ v(x)\right)) = |u(x) - v(x)|^p$ for the L^p case $(p = 1, 2)$, and that the regularization parameter β is some positive constant. In this paper, we propose a fast algorithm which computes an exact minimizer of problem 1. It relies on reformulating this problem into independent binary MRFs attached to each level set of an image. Exact minimization is performed thanks to a minimum cost cut algorithm.

The rest of this paper is organized as follows. In section 2 we map the original problem 1 into independent binary Markov Random Field optimizations. In section 3, a fast algorithm based on graph cuts is presented. In section 4 we shed new lights on tv minimization under the L^1-norm as fidelity term. Finally we draw some conclusions in section 5.

2 Formulation using Level Sets and MRF

For the rest of this paper we assume that u takes values in the discrete set $[0, L-1]$ and is defined on a discrete lattice S. We denote by u_s the value of the image u at the site $s \in S$. Let us decompose an image into its level sets using the decomposition principle [11]. It corresponds to considering the thresholding image u^λ where $u_s^\lambda = \mathbb{1}_{u_s \leq \lambda}$. One can reconstruct the original image from its level sets using $u_s = \min\{\lambda, u_s^\lambda = 1\}$.

2.1 Reformulation into binary MRFs

The coarea formula states that for any function u which belongs to the space of bounded variation [9] one has $tv(u) = \int_{\mathbb{R}} P(u^\lambda)\, d\lambda$ almost surely. In the discrete case, we write $tv(u) = \sum_{\lambda=0}^{L-2} P(u^\lambda)$, where $P(u^\lambda)$ is the perimeter of u^λ (notice that $u_s^{L-1} = 1$ for every $s \in S$, which explains the previous summation up to $L-2$.) Let us define the neighboring relationship between two sites s and t as $s \sim t$. The associated cliques of order two are noted as (s,t). This enables to estimate the perimeter using the approach proposed in [14]. Thus we have $tv(u) = \sum_{\lambda=0}^{L-2} \sum_{(s,t)} w_{st}\ |u_s^\lambda - u_t^\lambda|$, where w_{st} is set to 0.26 and 0.19 for the four- and eight- connected neighborhood, respectively.

Proposition 1 *The discrete version of the energy $E(u)$ rewrites as*

$$E(u) = \sum_{\lambda=0}^{L-2} E^\lambda(u^\lambda) + C \ , \ where \tag{2}$$

$$E^\lambda(u^\lambda) = \beta \left[\sum_{(s,t)} w_{st}\ ((1-2u_t^\lambda)\ u_s^\lambda + u_t^\lambda) \right] + \sum_{s \in \Omega} (g_s(\lambda+1) - g_s(\lambda))(1-u_s^\lambda) \tag{3}$$

$$g_s(x) = f(x,\ v_s)\ \ \forall s \in S \ \ and \ \ C = \sum_{s \in \Omega} g_s(0)$$

Proof: Using the following property for binary variables a, b: $|a-b| = a+b-2ab$, and starting from the previous equality obtained from the coarea

formula we have $tv(u) = \sum_{\lambda=0}^{L-2} \sum_{(s,t)} w_{st} \left((1 - 2u_t^\lambda)\, u_s^\lambda + u_t^\lambda \right)$. Moreover the following decomposition property holds for any function g :

$$\forall k \in [0, L-1]\ g(k) = \sum_{\lambda=0}^{k-1} ((g(\lambda+1) - g(\lambda)) + g(0) = \sum_{\lambda=0}^{L-2} (g(\lambda+1) - g(\lambda))\, \mathbb{1}_{\lambda<k} + g(0)$$

(note that this formula is coherent for both $k = 0$ and $k = L - 1$). Thus, by defining $g_s(u_s) = f(u_s,\ v_s)$ and since $\mathbb{1}_{\lambda<u_s} = 1 - u_s^\lambda$, we have

$$f(u_s,\ v_s) = g_s(u_s) = \sum_{\lambda=0}^{L-2} (g_s(\lambda+1) - g_s(\lambda))\, (1 - u_s^\lambda) + g_s(0) \ .$$

This concludes the proof. □

Note that each $E^\lambda(u^\lambda)$ is a binary MRF with an Ising prior model. To minimize $E(.)$ one can minimize all $E^\lambda(.)$ independently. Thus we get a family $\{\hat{u}^\lambda\}$ which are respectively minimizers of $E^\lambda(.)$. Clearly the summation will be minimized and thus we have a minimizer of $E(.)$ provided this family is monotone:

$$\hat{u}^\lambda \leq \hat{u}^\mu\ \forall \lambda < \mu \quad . \tag{4}$$

If this property holds then the optimal solution is given by [11] : $\hat{u}_s = \min\{\lambda, \hat{u}_s^\lambda = 1\}\ \forall s$. If property 4 does not hold, then the family $\{u^\lambda\}$ is not a function.

2.2 A Lemma based on coupled Markov Chains

Since the MRF posterior energy is decomposable into levels, it is useful to define the "local neighborhood configurations": $N_s = \{u_t\}_{t\sim s}$ and $N_s^\lambda = \{u_t^\lambda\}_{t\sim s}\ \forall \lambda \in [0, L-2]$. In [8] the following lemma was established:

Lemma 1 *If the local conditional posterior energy at each site s writes as*

$$E(u_s \mid N_s,\ v_s) = \sum_{\lambda=0}^{L-2} (\ \Delta\phi_s(\lambda)\ u_s^\lambda\ +\ \chi_s(\lambda)\ ) \tag{5}$$

where $\Delta\phi_s(\lambda)$ is a non-increasing *function of λ and $\chi_s(\lambda)$ does not depend on u_s^λ, then one can exhibit a "coupled" stochastic algorithm minimizing each total posterior energy $E^\lambda(u^\lambda)$ while preserving the monotone condition: $\forall s\ ,\ u_s^\lambda \nearrow$ with λ .*

In other words, given a binary solution $u^\star$ to the problem E^k, there exists at least one solution $\hat{u}$ to the problem E^l such that $u^\star \leq \hat{u}\ \forall k \leq l$. The proof of the Lemma relies on coupled Markov chains [20].

Proof: We endow the space of binary configurations by the following order : $u \leq v$ iff $u_s \leq v_s\ \forall s \in \Omega$. From the decomposition (5) the local conditional posterior energy at level value λ is $E(u_s^\lambda \mid N_s^\lambda,\ v_s) = \Delta\phi_s(\lambda)\ u_s^\lambda + \chi_s(\lambda)$. Thus let us define the following Gibbs local conditional posterior probability:

$$P_s(\lambda) = P(u_s^\lambda = 1 \mid N_s^\lambda,\ v_s) \quad = \quad \frac{\exp -\Delta\phi_s(\lambda)}{1 + \exp -\Delta\phi_s(\lambda)} = \frac{1}{1 + \exp \Delta\phi_s(\lambda)} \quad .(6)$$

With the conditions of the Lemma 1, this latter expression is clearly a monotone non-decreasing function of λ.

Let us now design a "coupled" Gibbs sampler for the $L-1$ binary images in the following sense: first consider a visiting order of the sites (tour). When a site s is visited, pick up a *single* random number ρ_s uniformly distributed in $[0, 1]$. Then, for each value of λ, assign: $u_s^\lambda = 1$ if $0 \leq \rho_s \leq P_s(\lambda)$ or else $u_s^\lambda = 0$ (this is the usual way to draw a binary value according to its probability, except that we use here the same random number ρ_s for all the $L-1$ binary images.) From the non-decreasing monotony of (6) it is seen that the set of assigned binary values at site s satisfies $u_s^\lambda = 1 \Rightarrow u_s^\mu = 1\ \forall \mu > \lambda$. The monotone property $u^\lambda \leq u^\mu\ \forall\ \lambda < \mu$ is thus preserved. Clearly, this property also extends to a series of $L-1$ coupled Gibbs samplers having *the same* positive temperature T when visiting a given site s: it suffices to replace $\Delta\phi_s(\lambda)$ by $\Delta\phi_s(\lambda)\ /\ T$ in (6). Hence, this property also holds for a series of $L-1$ coupled Simulated Annealing algorithms [10] where a *single* temperature T boils down to 0 (either after each visited site s or at the beginning of each tour [25] .) □

It must be noticed that our Lemma gives a *sufficient* condition for the simultaneous, "level-by-level independent" minimization of posterior energies while preserving the monotone property. We shall now prove the following property:

Lemma 2 *The requirements stated by Lemma 1 are equivalent to these:*
all conditional energies $E(u_s \mid N_s,\ v_s)$ are convex *functions of grey level $u_s \in [0, L-1]$, for any neighborhood configuration and local observed data.*

Proof: Since from (2) the total energy is "decomposable" on the levels, so are the local conditional energies: $E(u_s \mid N_s,\ \ v_s) = \sum_{\lambda=0}^{L-2} E^\lambda(u_s^\lambda \mid N_s^\lambda,\ \ v_s)$.

Besides, since the local conditional posterior energy at level λ is a function of binary variable u_s^λ, it satisfies:

$$E^\lambda(u_s^\lambda \mid N_s^\lambda,\ v_s) - E^\lambda(u_s^\lambda = 0 \mid N_s^\lambda,\ v_s) = \left(E^\lambda(u_s^\lambda = 1 \mid N_s^\lambda,\ v_s) - E^\lambda(u_s^\lambda = 0 \mid N_s^\lambda,\ v_s)\right)\ u_s^\lambda$$

which yields by identification with (5):

$$\Delta\phi_s(\lambda) = E^\lambda(u_s^\lambda = 1 \mid N_s^\lambda,\ v_s) - E^\lambda(u_s^\lambda = 0 \mid N_s^\lambda,\ v_s)$$

Now, in the transition $\lambda \to \lambda + 1$, only the following level variable does change: $u_s^\lambda = 1 \to u_s^\lambda = 0$. From the decomposition of conditional energies on levels, this means that only the level component $E^\lambda(u_s^\lambda \mid N_s^\lambda,\ v_s)$ does change and thus:

$$E(\lambda{+}1 \mid N_s,\, v_s) - E(\lambda \mid N_s,\, v_s) = E^\lambda(u_s^\lambda = 0 \mid N_s^\lambda,\, v_s) - E^\lambda(u_s^\lambda = 1 \mid N_s^\lambda,\, v_s) = -\Delta\phi_s(\lambda)$$

The monotone non-increasing condition on $\Delta\phi_s(\lambda)$ is thus equivalent to:

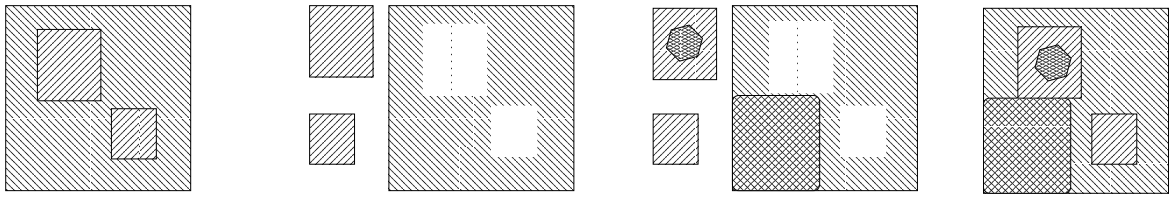
$E(\lambda{+}1 \mid N_s,\ v_s) - E(\lambda \mid N_s,\ v_s)$ is a *non-decreasing* function on $[0, L{-}2]$. □

Clearly both L^1 + TV and L^2 + TV models enjoy this convexity property and satisfy thus the conditions of application of Lemma 1.

3 Minimization Algorithm

Although the previous section proves that the monotone property holds, it does not provide an algorithm to compute a solution. Our algorithm makes use of the formulation shown in equation 2 which allows independent optimizations. A natural algorithm, presented in [8], is to optimize independently each MRF. This leads to an algorithm which performs $L-1$ optimizations on binary images whose sizes are the same as the original image.

However, one can both drastically save computations using a divide and conquer approach. Such an approach requires to decompose a problem into smaller ones, then to solve these sub-problems and to recombine the sub-solutions to get an optimal solution. Our algorithm takes benefit of the following. Suppose we minimize at some level λ. Then, for all pixels of the minimizer we know whether they are below or above λ. Thus it is useless to take into account pixels above λ for further optimizations which only allow pixels to be lower than λ. Obviously, the same holds for pixels which are below λ. Then, every connected component (it defines a partition of the image) of the minimizer can be optimized independently from each others. The latter corresponds to the decomposition of the problem into subproblems. Once minimizers of subproblems are computed,



(a) (b) (c) (d)

Figure 1: Illustration of our algorithm. The partition of the image after a minimization with respect to some level λ is shown on (a). The connected components of the image (a) are shown on (b): it corresponds to the decomposition of the problem into subproblems. Each subproblem is solved independently and the result is depicted in (c). Finally solutions of subproblems are recombined to yield the image (d).

Table 1: Time results (in seconds) with L^1 data fidelity term for different weighted term β. Two time results are presented: time for our algorithm and time for the algorithm presented in [8] inside parentheses.

Image	$\beta = 1$	$\beta = 2$	$\beta = 3$
Lena (256x256)	0.37(7.31)	0.54(14.52)	0.72(16.41)
Lena (512x512)	1.56(31.10)	2.24(53.36)	2.84(101.33)
Woman (522x232)	0.53(16.03)	0.77(20.34)	1.03(23.86)

they are recombined to yield an optimal solution. The recombination is straightforward since the decomposition was a partition. This process is depicted in Figure 1. A good choice to choose the threshold level λ is to use a dichotomic process. For instance, suppose the minimizer is a constant image, then our algorithm requires exactly $\log_2(L)$ (we suppose L is a power of two) binary optimizations to compute it. This is in contrast compared to the $L - 1$ required binary optimizations of the algorithm proposed in [8].

Optimization of a binary MRF can be performed exactly and efficiently using graph-cut techniques. It consists of building a graph such that its minimum cut gives an optimal labelling. We build the graph as proposed in [13]. Our implementation uses the minimum cut algorithm described in [3]. Time results (on a 3GHz Pentium IV) for our algorithm and the one presented in [8] are given in Table 1 for L^1 fidelity. Note how our algorithm outperforms the other one.

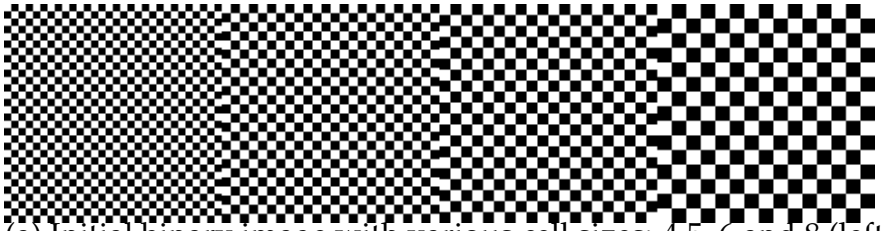
4 Total Variation with L^1 data fidelity

The use of total variation with L^1 data fidelity has already been studied in [2, 6, 15, 16]. However, the following is new as far as we know. Note that the Ising model fulfills the necessary condition provided that the interaction is attractive (i.e. β is non-negative) which is the case in our problems.
As a matter of fact, due to the equivalence of the Potts framework, the initial $L_1 + TV$ restoration model (assign $g_s(u_s) = |u_s - v_s| = \sum_{\lambda=0}^{L-2} |u_s^\lambda - v_s^\lambda|$ in (3)) is equivalent to an Ising model with constant magnetic field amplitude $B = 1/2$ and constant interaction coefficient $J = \beta/2$ over the whole range of levels.
It was shown, first semi-empirically [23] and then rigorously [18] that the 4-connnected chessboard model exhibits a phase transition property. Namely if the basic cell size A satisfies: $A \leq 4J/B = 4\beta$ then two ground states occur, corresponding to uniform binary images. In the opposite case, the unique ground state is the initial chessboard itself. In other words, and put in a rather "inexact" way, objects whose characteristic size is greater than 4β are conveniently restored, whereas smaller objects are lost in the "background". This property holds on the whole range of levels for the $L^1 + TV$ model (See Fig. 2).

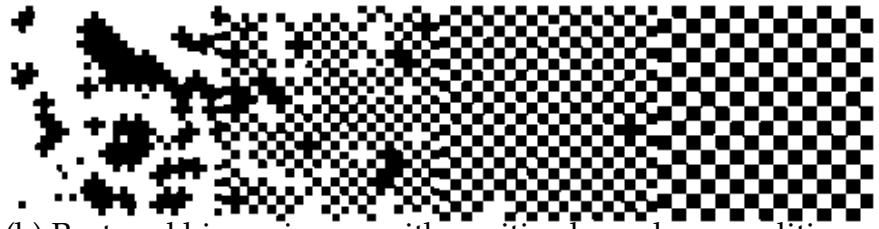
Moreover, it was shown in [7] that the continuous approach to this problem generates extra grey levels outside the initial grey level range, which is obviously not the case here. This happens because of the ϵ introduced in the numerical scheme to avoid division by zero. Figure 3 depicts some results on the image woman. Note how the contrast is well preserved and how level lines are simplified.

5 Conclusion

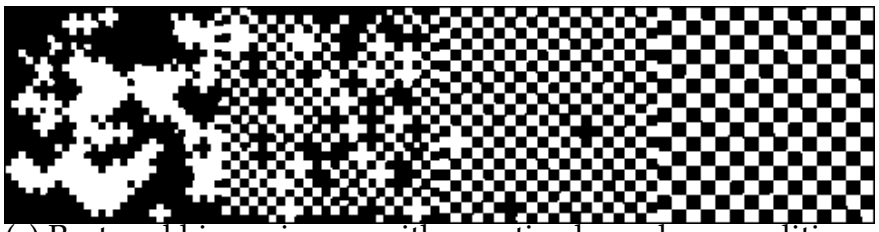
In this paper we have presented an algorithm which computes an exact solution for the minization of the total variation under a convex constraint. The method relies on the decomposition of the problem into binary ones using the level sets of an image. Moreover, this algorithm is quite fast. Comparison to other algorithms with respect to time complexity must be made. Extension of this method to other type of regularization is in progress.



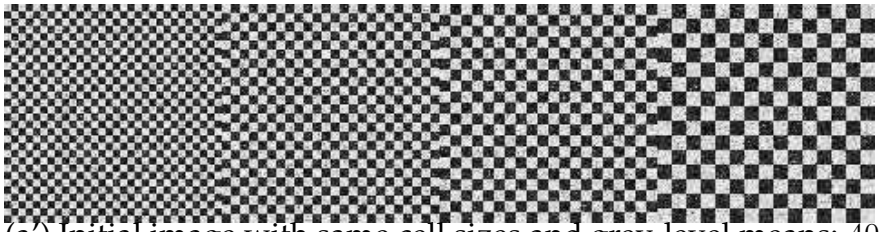
(a) Initial binary image with various cell sizes: 4,5, 6 and 8 (left→right).



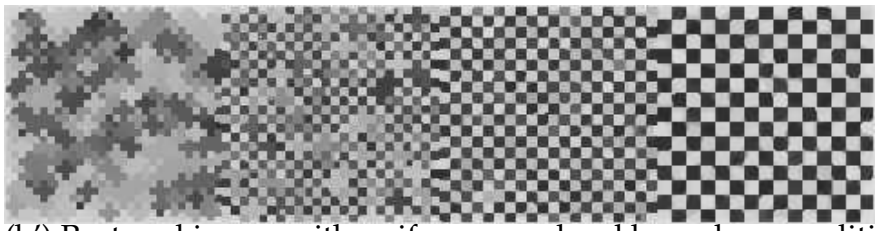
(b) Restored binary image with positive boundary conditions.



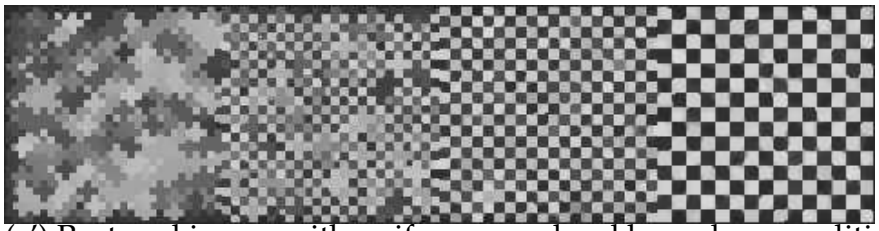
(c) Restored binary image with negative boundary conditions.



(a') Initial image with same cell sizes and grey-level means: 40 and 220.



(b') Restored image with uniform grey-level boundary conditions: 220.



(c') Restored image with uniform grey-level boundary conditions: 40.

Figure 2: Minimal energy configurations obtained by Simulated Annealing. Initial temperature $T_0 = 16$ with decreasing step = 0.98, $\beta = 1.5$ (4-connectivity).

References

[1] S. Alliney. An algorithm for the minimization of mixed l^1 and l^2 norms with application to bayesian estimation. *IEEE Signal Processing*, 42(3):618–627, 1994.

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

[3] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE PAMI*, 26(9):1124–1137, 2004.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE PAMI*, 23(11):1222–1239, 2001.

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

[6] T.F. Chan and S. Esedoğlu. Aspect of total variation regularized l^1 function approximation. Technical Report 7, UCLA, 2004.

[7] T.F. Chan, S. Esedoglu, and M. Nikolova. Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models. Technical report, UCLA, September 2004.

[8] J. Darbon and M. Sigelle. Exact Optimization of Discrete Constrained Total Variation Minimization Problems. In LNCS series vol. 3322, editor, *Tenth International Workshop on Combinatorial Image Analysis*, pages 540–549, 2004.

[9] L. Evans and R. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, 1992.

[10] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE PAMI*, 6(6):721–741, 1984.

[11] F. Guichard and J.M. Morel. Mathematical morphology "almost everywhere". In *Proceedings of ISMM*, pages 293–303. Csiro Publishing, April 2002.

[12] H. Ishikawa. Exact optimization for Markov random fields with priors. *IEEE PAMI*, 25(10):1333–1336, November 2003.

[13] V. Kolmogorov and R. Zabih. What energy can be minimized via graph cuts? *IEEE PAMI*, 26(2):147–159, 2004.

[14] H.T. Nguyen, M. Worring, and R. van den Boomgaard. Watersnakes: Energy-driven watershed segmentation. *IEEE PAMI*, 23(3):330–342, 2003.

[15] M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. *SIAM J. Num. Anal.*, 40(3):965–994, 2002.

[16] M. Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20:99–120, 2004.

[17] S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the $\mathbf{H}^{-1}norm$. *J. Mult. Model. and Simul.*, 1(3), 2003.

[18] E. Pechersky, A. Maruani, and M. Sigelle. On Gibbs Fields in Image Processing. *Markov Processes and Related Fields*, 1(3):419–442, 1995.

[19] I. Pollak, A.S. Willsky, and Y. Huang. Nonlinear evolution equations as fast and exact solvers of estimation problems. *IEEE Signal Processing*, 53(2):484–498, 2005.

[20] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and statistical mechanics. *Random Structures and Algorithms*, 9(1):223–252, 1996.

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

[22] K. Sauer and C. Bouman. Bayesian estimation of transmission tomograms using segmentation based optimization. *IEEE Nuclear Science*, 39(4):1144–1152, 1992.

[23] M. Sigelle. *Champs de Markov en Traitement d'Images et Modèles de la Physique Statistique : Application la Relaxation d'Images de Classification*. PhD thesis, ENST http://www.tsi.enst.fr/∼sigelle/tsi-these.html, 1993.

[24] C. Vogel and M. Oman. Iterative method for total variation denoising. *SIAM J. Sci. Comput.*, 17:227–238, 1996.

[25] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Applications of mathematics. Springer-Verlag, 2003.

Figure 3: Minimizers of TV with L^1 fidelity. From left to right: Original image, then minimizers for $\beta = 1$, $\beta = 2.1$, $\beta = 3$ on the the first row. Finally, some level lines of the original image and of the minimizers (in the same order) on the second row. Only level lines multiples of 10 are displayed.