Exact Optimization of Discrete Constrained Total Variation Minimization Problems

Jérôme Darbon[1,2] and Marc Sigelle[2]

[1] EPITA,
14-16 rue Voltaire, F-94276 Le Kremlin-Bicêtre, France,
jerome.darbon@{lrde.epita.fr, enst.fr}
[2] ENST,
46 rue Barrault, F-75013 Paris, France
marc.sigelle@enst.fr

Abstract. This paper deals with the total variation minimization problem when the fidelity is either the L^2-norm or the L^1-norm. We propose an algorithm which computes the exact solution of these two problems after discretization. Our method relies on the decomposition of an image into its level sets. It maps the original problems into independent binary Markov Random Field optimization problems associated to each level set. Exact solutions of these binary problems are found thanks to minimum-cut techniques. We prove that these binary solutions are increasing and thus allow to reconstruct the solution of the original problems.

1 Introduction

Image reconstruction and deconvolution methods are often based on the minimization of the constrained total variation [2, 22, 26, 27, 29] of an image u defined on Ω. These problems have minimizers in the space of functions of bounded variation [12] which allows for discontinuities and thus preserve edges and sharp boundaries. Suppose u is defined on a rectangle Ω of $\mathbb{R}^2$. Then the total variation of u, $TV(u)$, is defined as follows:

$$TV(u) = \int_{\Omega} |\nabla u| \ ,$$

where the gradient of u is taken in the distributional sense. A classical way to minimize the total variation is achieved by a gradient descent which yields the following evolution equation:

$$\frac{\partial u}{\partial t} = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \ .$$

The last term corresponds to the curvature of u. In order to avoid division by zero, a classical approximation is to replace $|\nabla u|$ by $\sqrt{|\nabla u|^2 + \epsilon}$, as discussed in [1, 30]. However, this scheme tends to smooth discontinuities and although it converges towards the solution when ϵ tends to 0, it does not provide an exact

solution. Other formulations of TV minimization using duality are presented in [8, 9]. A fast algorithm which converges towards the solution can be dervied from this formulation. In [7], a fast approximation minimization algorithm for Markov Random Field (MRF) is presented. It relies on minimum cost cut and the result is a local minimum.

In [23], a fast algorithm to compute the exact solution in 1D for the TV minimization problem subject to the L^2 constraint is presented. However, the algorithm does not scale to higher dimensions. In 1D, one can find an exact solution using dynamic programming [4], provided that the label state is discrete. The complexity of such a method is $\Theta(N^2|\Omega|)$, where N and $|\Omega|$ are the cardinality of the label state and the number of pixels in the discrete domain Ω, respectively. In [16], Ishikawa presents an algorithm to find the exact solution for Markov Random Field with convex priors in a polynomial time.

In this paper, we focus on TV minimization with L^1 or L^2 fidelity. Thus, we are interested in minimizing the following functionals:

$$E_\alpha(u,\beta) = \int_\Omega |u(x) - v(x)|^\alpha \, dx + \beta \int_\Omega |\nabla u| \quad ,$$

where $\alpha \in \{1, 2\}$ and $\beta \geq 0$. The use of the L^1 fidelity has already been studied in [3, 10, 21, 20]. Very good results for image denoising are reported in [21]. Our main contribution is an exact optimization of a discretization of the two functionals $E_\alpha(., \beta)$. It relies on reformulating the original problem into several independent binary problems which are expressed through the Markov random field framework. The reformulation is based on the decomposition of a function into its level sets.

The rest of this paper is as follows. The decomposition of the considered problems into independent binary problems is described in section 2. In section 3, reconstruction of the solution from solutions of the binary problems is shown. Minimization algorithm and results are presented in section 4. Finally we draw some conclusions in section 5.

2 Formulation through Level Sets

In this section, we show that minimization of the TV minimization problem with L^1 or L^2 fidelity can be decomposed into the minimization of independent binary problems. For each level $\lambda \in [0, N-1]$, we consider the thresholded images u^λ of an image: $u^\lambda = \mathbb{1}_{u \leq \lambda}$. Note that this decomposition is sufficient to reconstruct the gray-level image: $u(x) = \min\{\lambda, u^\lambda(x) = 1\}$.

2.1 Coarea Formula

For any function $u \in BV(\Omega)$, the Coarea formula [12] gives

$$TV(u) = \int_{\mathbb{R}} P(u^\lambda) \mathrm{d}\lambda \quad ,$$

for almost all λ and where $P(u^\lambda)$ is the perimeter of u^λ. We estimate the perimeter using pairs of neighboring pixels:

$$TV(u) = \sum_{\lambda=0}^{N-1} \sum_{s \sim t} R_{s,t}(u_s, v_s, \lambda) \ , \tag{1}$$

where $s \sim t$ denotes neighboring pixels and

$$R_{s,t}(u_s, v_s, \lambda) = w_{s,t} |u_s^\lambda - u_t^\lambda| \ .$$

For our experiments we use two different contour length estimators. The first one consists in considering only the four-connected neighborhood and setting $w_{s \sim t}$ to 1. The second one, as proposed in [19], sets $w_{s,t}$ to 0.26 and 0.19 for the four and eight connected neighborhood respectively. Note that the latter estimation is not accurate for small regions.

2.2 Expressing L^1 through Level Sets

We reformulate L^1 fidelity into level sets. First we decompose the domain into the following two disjoint sets $\{s : u_s < v_s\}$ and $\{s : u_s > v_s\}$. This yields

$$\sum_{s \in \Omega} |u_s - v_s| = \sum_{u_s < v_s} |v_s - u_s| + \sum_{u_s > v_s} |u_s - v_s| \ .$$

Then we rewrite the absolute norm as a sum:

$$\sum_{s \in \Omega} |u_s - v_s| = \sum_{u_s < v_s} \sum_{\lambda = u_s}^{v_s - 1} 1 + \sum_{u_s > v_s} \sum_{\lambda = v_s}^{u_s - 1} 1 \ .$$

Using the indicator functions of the disjoint sets, we have

$$\sum_{s \in \Omega} |u_s - v_s| = \sum_{s \in \Omega} \sum_{\lambda=0}^{N-1} \mathbb{1}_{u_s < v_s}(s) \mathbb{1}_{[u_s, v_s - 1]}(\lambda) + \mathbb{1}_{u_s < v_s}(s) \mathbb{1}_{[v_s, u_s - 1]}(\lambda) \ .$$

One can easily check that the term inside summations is either 1 or 0 and is equal to $|u_s^\lambda - v_s^\lambda|$. Consequently by interchanging the order of summation, it follows that

$$\sum_{s \in \Omega} |u_s - v_s| = \sum_{\lambda=0}^{N-1} \sum_{s \in \Omega} D_1(u_s, v_s, \lambda) \ , \tag{2}$$

where

$$D_1(x, y, \lambda) = |x^\lambda - y^\lambda| \ .$$

Note that this formulation shows that the L^1-norm treats level sets of the image u independently of their associated gray-levels. This can be seen as adopting a geometrical point of view.

2.3 Expressing L^2 through Level Sets

The same approach is used for the decomposition of L^2 into level sets. However, contrary to the L^1 norm, the decomposition cannot be independent of its gray-levels. We begin with separating the sum according to the two previous disjoint sets:

$$\sum_{s\in\Omega}(u_s - v_s)^2 = \sum_{u_s<v_s}(u_s - v_s)^2 + \sum_{u_s>v_s}(u_s - v_s)^2.$$

Using the formula $\sum_{k=1}^{M}(2k-1) = M^2$, we have:

$$\sum_{s\in\Omega}(u_s - v_s)^2 = \sum_{u_s<v_s}\sum_{k=1}^{v_s-u_s}(2k-1) + \sum_{u_s>v_s}\sum_{l=1}^{u_s-v_s}(2l-1).$$

For the first sum we make the following change of variable $k \leftarrow v_s - \lambda$, while we do $l \leftarrow \lambda - v_s + 1$ for the second one. It leads to:

$$\sum_{s\in\Omega}(u_s - v_s)^2 = \sum_{u_s<v_s}\sum_{\lambda=u_s}^{v_s-1}(2(v_s-\lambda)-1) + \sum_{u_s>v_s}\sum_{\lambda=v_s}^{u_s-1}(2(\lambda-v_s)+1).$$

Then, we introduce the characteristic function of the disjoint sets and interchange the order of summation. Moreover, since $2v_s^\lambda - 1 = 1$ iff $v_s \leq \lambda$ and -1 otherwise, it follows that:

$$\sum_{s\in\Omega}(u_s - v_s)^2 = \sum_{\lambda=0}^{N-1}\sum_{s\in\Omega}(\mathbb{1}_{u<v}(s) + \mathbb{1}_{u>v}(s))\left(2(|v_s-\lambda| + 2v_s^\lambda - 1)\right) \quad .$$

Finally notice that $\mathbb{1}_{u<v}(s) + \mathbb{1}_{u>v}(s) = |u_s^\lambda - v_s^\lambda|$. Consequently, we have:

$$\sum_{s\in\Omega}(u_s - v_s)^2 = \sum_{\lambda=0}^{N-1}\sum_{s\in\Omega}D_2(u_s, v_s, \lambda) \tag{3}$$

where

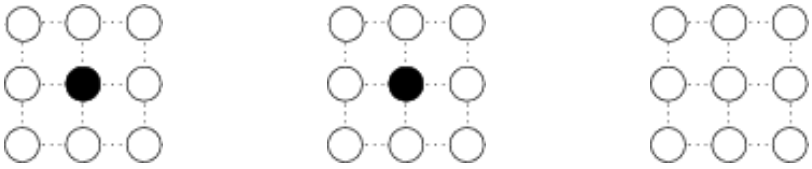
$$D_2(x, y, \lambda) = |x^\lambda - y^\lambda|\left(2(|y-\lambda| + 2y^\lambda - 1)\right)$$

This formulation shows that L^2 can be decomposed into level sets where their associated gray-levels are taken into account.

2.4 Independent Optimizations

Finally, both energies can be re-written as follows:

$$E_\alpha(u,\beta) = \sum_{\lambda=0}^{N-1}\left(\sum_{s\in\Omega}D_\alpha(u_s, v_s, \lambda) + \beta\sum_{s\sim t}w_{s,t}|u_s^\lambda - u_t^\lambda|\right) = \sum_{\lambda=0}^{N-1}E_\alpha^\lambda(u,\beta) \quad .$$



(a) (b) (c)

Fig. 1. Since $E_1(.,\beta)$ is not strictly convex, minimizers can be non-unique. The original image is depicted in (a) where 4-connectivity is considered. Black and white circles refer to sites whose value is 0 and 1, respectively. If $\beta = 0.25$ then there are two minimizers depicted in (b) and (c), whose associated energy is 1.

Note that the term $E_\alpha^\lambda(u,\beta)$ is a Markov Random Field which only involves binary variables and pairwise interactions. The prior is an Ising model [31]. Thus for each λ, one deals with a binary MRF.

Now suppose that for each λ, we independently find the best binary configuration $\hat{u}^\lambda$ which minimizes the energy of the MRF. Clearly, the summation will be minimized. Thus we will find a minimizer for $E_\alpha(.,\beta)$ provided that the following property of monotony holds for binary minimizers:

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

Indeed, if this property holds, then the minimizer $\hat{u}$ of $E_\alpha(.,\beta)$ is given by [15]

$$\hat{u}_s = \min\{\lambda, \hat{u}_s^\lambda = 1\} \ \forall s \ .$$

The proof of the monotone property is given in the next section.

3 Reconstruction of the solution

In this section, we prove the monotone property defined by (4). However, since $E_1(.,\beta)$ is not strictly convex, it leads to non-unique minimizers in general. Such a situation is depicted in figure 1. The monotone property can be violated in that case. However, we claim the following result:

Lemma
If each local conditional posterior energy can be written up to a constant, as

$$\mathcal{W}(u_s \mid \{u_t\},\ v_s) = \sum_{\lambda=0}^{N-1} \phi_s(\lambda)\ u_s^\lambda \tag{5}$$

where $\phi_s(\lambda)$ is a non-increasing function of λ, then one can exhibit a "coupled" stochastic algorithm minimizing each associated total posterior energy on the N binary images u^λ, 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_α, there exists at least one solution $\hat{u}$ to the problem E^l_α such that $u^\star \leq \hat{u}\ \forall k \leq l$.
First, let us show that TV-like regularization and attachment to data energies both feature property (5), so will do their sum and thus the total posterior energy. Since for any binary variables a, b we have: $|a - b| = a + b - 2ab$, and considering (1), this yields

$$|u_s - u_t| = \sum_{\lambda=0}^{N-1} |u_s^\lambda - u_t^\lambda| = \sum_{\lambda=0}^{N-1} (1 - 2u_t^\lambda)\ u_s^\lambda\ + C$$

where "constant" $C = \sum_{\lambda=0}^{N-1} u_t^\lambda$ only depends on the $\{u_t^\lambda\}$. Thus, we have $\phi_s(\lambda) = 1 - 2u_t^\lambda$, which is by essence a non-increasing function of λ.
Similarly starting from (2) for L^1 attachment to data term:

$$|u_s - v_s| = \sum_{\lambda=0}^{N-1} |u_s^\lambda - v_s^\lambda| = \sum_{\lambda=0}^{N-1} (1 - 2v_s^\lambda)\ u_s^\lambda\ + C' \quad , \quad C' = \sum_{\lambda=0}^{N-1} v_s^\lambda$$

The approach for the L^2 relies on the same method. Recall from (3) that

$$(u_s - v_s)^2 = \sum_{\lambda=0}^{N-1} \sum_{\lambda=0}^{N-1} \phi_s(\lambda) + C'' \ ,$$

where

$$\phi_s(\lambda) = (1 - 2v_s^\lambda)\left(2(|v_s - \lambda| + 2v_s^\lambda - 1)\right) u_s^\lambda \ .$$

One can easily check that $\phi_s(\lambda)$ fulfills our requirement. Here also the constant C'' does not depend on $\{u_s^\lambda\}$(only on v_s).

Thus we have shown that TV regularization with either L^1 and L^2 attachment to data terms both follow the conditions of our Lemma.
In both cases, the proof of the Lemma relies on coupled Markov chains [24, 11].

3.1 Coupled Markov Chains

We endow the space of binary configurations by the following order : $u \leq v$ iff $u_s \leq v_s$ for all $s \in \Omega$.

From the decomposition (5) the local conditional posterior energy at level value λ is $\phi_s(\lambda)\ u_s^\lambda$. Thus the related Gibbs local conditional posterior probability is

$$P(u_s^\lambda = 1 \mid \{u_t^\lambda\},\ v_s^\lambda) = \frac{\exp -\phi_s(\lambda)}{1 + \exp -\phi_s(\lambda)} = \frac{1}{1 + \exp \phi_s(\lambda)} \ . \tag{6}$$

With the conditions of the Lemma, this latter expression is clearly a monotone non-decreasing function of λ.
Let us now design a "coupled" Gibbs sampler for the N 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 uniformly distributed in $[0, 1]$. Then, for each value of λ, assign $u_s^\lambda = 1$ if $\rho \leq P(u_s^\lambda = 1 \mid \{u_t^\lambda\},\ v_s^\lambda)$ 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 for all the N 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 N coupled Gibbs samplers having **the same** positive temperature T when visiting a given site s: it suffices to replace $\phi_s(\lambda)$ by $\phi_s(\lambda)$ / T in (6). Hence, this property also holds for a series of N coupled Simulated Annealing algorithms [13] where a **single** temperature T boils down to 0 (either after each visited site s or at the beginning of each tour [31] .) This concludes the proof.

Several points should be emphasized here:

- first, the coupled monotony-preserving Gibbs/Metropolis samplers described in [24] relate to the same Markov Random Field but for various initial conditions, while here, our N coupled Gibbs samplers relate to N different posterior MRF's (one for each level λ).

- second, it must also 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.

Although we have proved the monotone property, it does not provide an algorithm to compute the solution. Indeed, using a Simulated Annealing process, one knows that it converges to the solution but has no stopping criteria.

4 Computations and Experiments

In this section, we describe our algorithm to exactly minimize each binary Markov Random Field. It relies on graph-cuts.We also present some experiments.

4.1 Minimum-cut based minimization

Greig et al. [14] were the first ones to propose an exact maximum *a posteriori* estimator for binary MRF. It is based on constructing a graph such that its minimum cut gives an optimal labelling. Since this seminal work, other graph constructions were proposed to solve some non-binary problems exactly ([25], [16]) or approximately ([7]). in [17], Kolmogorov and Zabih propose a necessary condition for any binary function to be minimized via minimum cuts along with a graph construction for this class of functions. 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.

For each level we construct the graph as proposed in [17] and compute a minimum cost cut. However, since uniqueness cannot be assured with L^1 fidelity, the algorithm returns one of the optimal configurations. Since these minimizations

are independently performed, the monotone property can be violated. In order to reconstruct the solution, one flips every pixel where this property is violated. This flipping process also gives an optimal labelling since it does not change the energy.

To compute the minimum cut, we used the algorithm described in [6]. Although its theoretical complexity is worse than others, it outperforms them in practice for graphs used in computer vision problems. For our binary problems, this algorithm gives near-linear performance with respect to the number of pixels $|\Omega|$. Since we have to compute N cuts, the complexity of our algorithm is near-linear both with respect to the number of pixels and with the number of labels. In [6], the authors study the method of Ishikawa [16] which also computes an exact solution. Experiments show that this method is near-linear with respect to the size of the image and near-quadratic with respect to the number of labels. Thus, compared to this method, our algorithm is faster by a factor of N. Time results (in seconds, 1.6GHz Pentium IV) for our method are presented in table 1 for L^1 fidelity applied to the image "hand" depicted in Fig 4.

Table 1. Time results (in seconds) with L^1 fidelity for the image "hand"

Size	Connectivity	Time
151x121	8	7.58
343x243	8	30.56
151x121	4	4.97
343x243	4	21.02

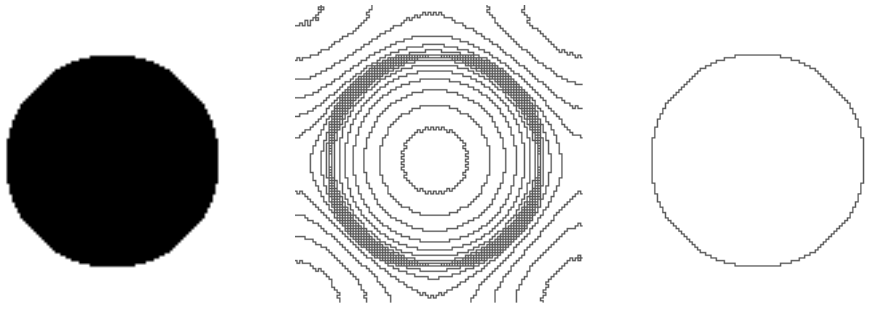
Note that the proposed algorithm does not benefit from the monotone property. If a binary solution is computed for a level, then every pixel which equals 1 will be at 1 for upper levels. Thus, one can prevent them from being in the graph built for upper levels. This would automatically guaranty the monotone property and save computations. Such a method will be studied in the future.

4.2 Experiments

For these experiments, we always use the 8-connectivity. In [28] the authors give exact and analytic solutions for TV minimization with L^2 attachment for radial symmetric functions. For instance, if the observed image is a circle then the solution is a circle with the same radius : only its gray-levels change. We verified that our algorithm produced the exact solution and compared it with the classical gradient descent algorithm. Recall that this latter algorithm needs $|\nabla u|$ to be regularized because of non-differentiability : $|\nabla u| = \sqrt{|\nabla u|^2 + \epsilon}$. The evolution equation is

$$\frac{\partial u}{\partial t} = \beta \mathrm{div}\left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \epsilon}}\right) + 2(v - u) \ .$$

We set $\epsilon = 1$ for this experiment. Figure 2 depicts the level-lines of the solutions for the two different algorithms. Note how many level lines are created by the gradient descent algorithms.



(a) (b) (c)

Fig. 2. Minimizers of TV under L^2 constraints ($\beta = 1$). The original is depicted in (a). The level lines resulting from the gradient descent algorithm is presented in (b). The level lines of the exact solution, computed using our algorithm, is depicted in (c).

Total variation is very well known for its high performance in image restoration. Figure 3 depicts a cartoon image and its noisy version corrupted by an additive Gaussian noise ($\sigma = 30$). It also presents the results of the restoration using the gradient descent method and our algorithm. Although the results visually look the same, the exact solution provides a much better result in terms of level lines. Note how corners of the squares are smoothed. This is predicted by the theory [5, 18] which states that a square cannot arise as a solution.

Results of the regularization using the L^1-norm as fidelity are depicted in figure 4. As one can see, the more β is high the more small structures are removed while the contrast is preserved.

5 Conclusion

In this paper we have presented an algorithm to compute the exact solution of the discrete TV-based restoration problem when fidelity is the L^1 or L^2 norm. It relies on the decomposition of the problem into binary ones thanks to a level set formulation. It allows for an algorithm whose complexity is near-linear both with respect to the image size and the number of labels.

Extension of this method to other types of fidelity is in progress. In particular, one can decompose any fidelity function f as follows:

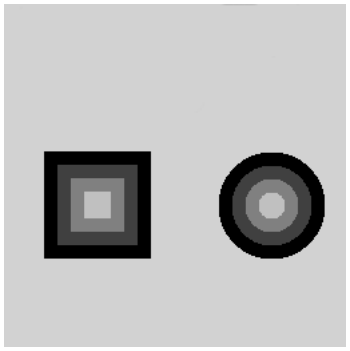
$$f(u_s) = \sum_{\lambda=0}^{N-1} (f(\lambda+1) - f(\lambda))\ \mathbb{1}_{\lambda<u_s}\ +\ f(0)$$
$$= \sum_{\lambda=0}^{N-1} (f(\lambda+1) - f(\lambda))\ (1-u_s^\lambda)\ +\ f(0)\ .$$

It follows that the condition stated by our Lemma is equivalent to the fact that each local conditional posterior energy is a **convex** function on the integer set $[0, N-1]$, since its finite difference is increasing. We are currently working on a less restrictive condition. Moreover, applying this method to regularization terms different from TV has to be considered. Finally a faster minimization algorithm which takes into account the monotone property is under study. Comparisons with other exact minimization algorithms must also be made.

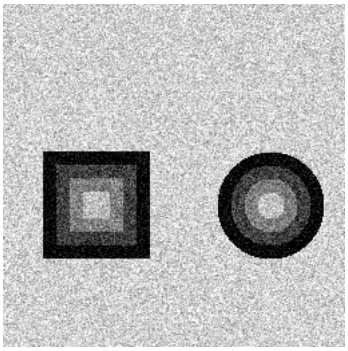
References

1. A. Acart and C. Vogel. Analysis of bounded variation penalty methods of ill-posed problems. *Inverse Problem*, pages 1217–1229, 1994.
2. S. Alliney. An algorithm for the minimization of mixed l^1 and l^2 norms with application to bayesian estimation. *IEEE Transactions on Signal Processing*, 42(3):618–627, 1994.
3. S. Alliney. A property of the minimum vectors of a regularizing functional defined by means of the absolute norm. *IEEE Transactions on Signal Processing*, 45(4):913–917, 1997.
4. A. Amini, T. Weymouth, and R. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 12(9):855–867, 1990.
5. S. Durand ans F. Malgouyres abd B. Rougé. Image deblurring, spectrum interpolation and application to satellite imaging. *Control Optimization and Calculus of Varaition*, 5:445–477, 2000.
6. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 26(9):1124–1137, 2004.
7. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 23(11):1222–1239, 2001.
8. A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.
9. T. Chan, G.H. Golub, and P. Mulet. A nonlinear primaldual method dor total variation-based image restoration. *SIAM Journal on Scientific Computing*, 20(6):1964–1977, 1999.
10. T.F. Chan and S. Esedoğlu. Aspect of total variation regularized l^1 function approximation. Technical Report 7, UCLA, 2004.
11. P.M. Djurić, Y. Huang, and T. Ghirmai. Perfect sampling : A review and applications to signal processing. *IEEE Transactions on Signal Processing*, 50(2):345–256, 2002.

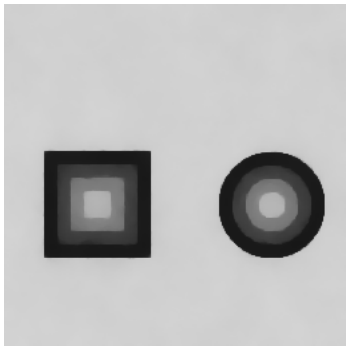
12. L. Evans and R. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Pres, 1992.
13. S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 6(6):721–741, 1984.
14. D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistics Society*, 51(2):271–279, 1989.
15. F. Guichard and J.M. Morel. Mathematical morphology "almost everywhere". In *Mathematical Morphology, Proceedings of the 6th International Symposium (ISMM)*, pages 293–303. Csiro Publishing, April 2002.
16. H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 25(10):1333–1336, November 2003.
17. V. Kolmogorov and R. Zabih. What energy can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Interaction*, 26(2):147–159, 2004.
18. Y. Meyer. Oscillating patterns in image processing and nonlinear evolution equations. *University Lecture Series*, 22, 2001.
19. H.T. Nguyen, M. Worring, and R. van den Boomgaard. Watersnakes: Energy-driven watershed segmentation. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 23(3):330–342, 2003.
20. M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers. *SIAM Journal on Numerical Analysis*, 40(3):965–994, 2002.
21. M. Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20:99–120, 2004.
22. S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the $\mathbf{H}^{-1}$*norm*. *SIAM J. Multiscale Modeling and Simulation*, 1(3), 2003.
23. I. Pollak, A.S. Willsky, and Y. Huang. Nonlinear evolution equations as fast and exact solvers of estimation problems. *to appear in IEEE Transactions on Signal Processing*, 2004.
24. J. G. Propp and D. B. Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1):223–252, 1996.
25. S. Roy. Stereo without epipolar lines: A maximum-flow formulation. *International Journal of Computer Vision*, 34(2):147–162, 1999.
26. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D.*, 60:259–268, 1992.
27. K. Sauer and C. Bouman. Bayesian estimation of transmission tomograms using segmentation based optimization. *IEEE Transactions on Nuclear Science*, 39(4):1144–1152, 1992.
28. D. Strong and T.F. Chan. Edge preserving and scale-dependent properties of total variation regularization. *Inverse Problem*, 19:165–187, 2003.
29. E. Tadmor, S. Nezzar, and L. Lese. A multiscale image representation using hierarchical (bv,l^2) decompositions. Technical report, UCLA, 2003.
30. C. Vogel and M. Oman. Iterative method for total variation denoising. *SIAM J. Sci. Comput.*, 17:227–238, 1996.
31. G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods. A Mathematical Introduction*. Applications of mathematics. Springer-Verlag, 2003.



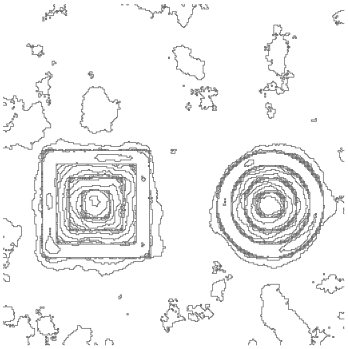
(a) Original image



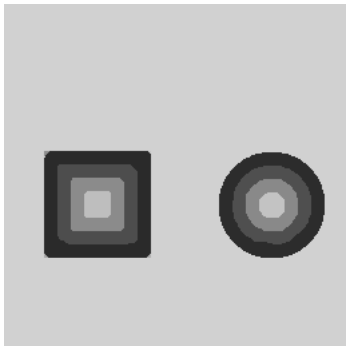
(b) Noisy image ($\sigma = 30$)



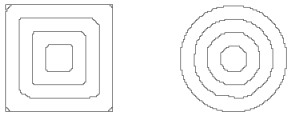
(c) Gradient descent restoration



(d) Some level lines of (c)

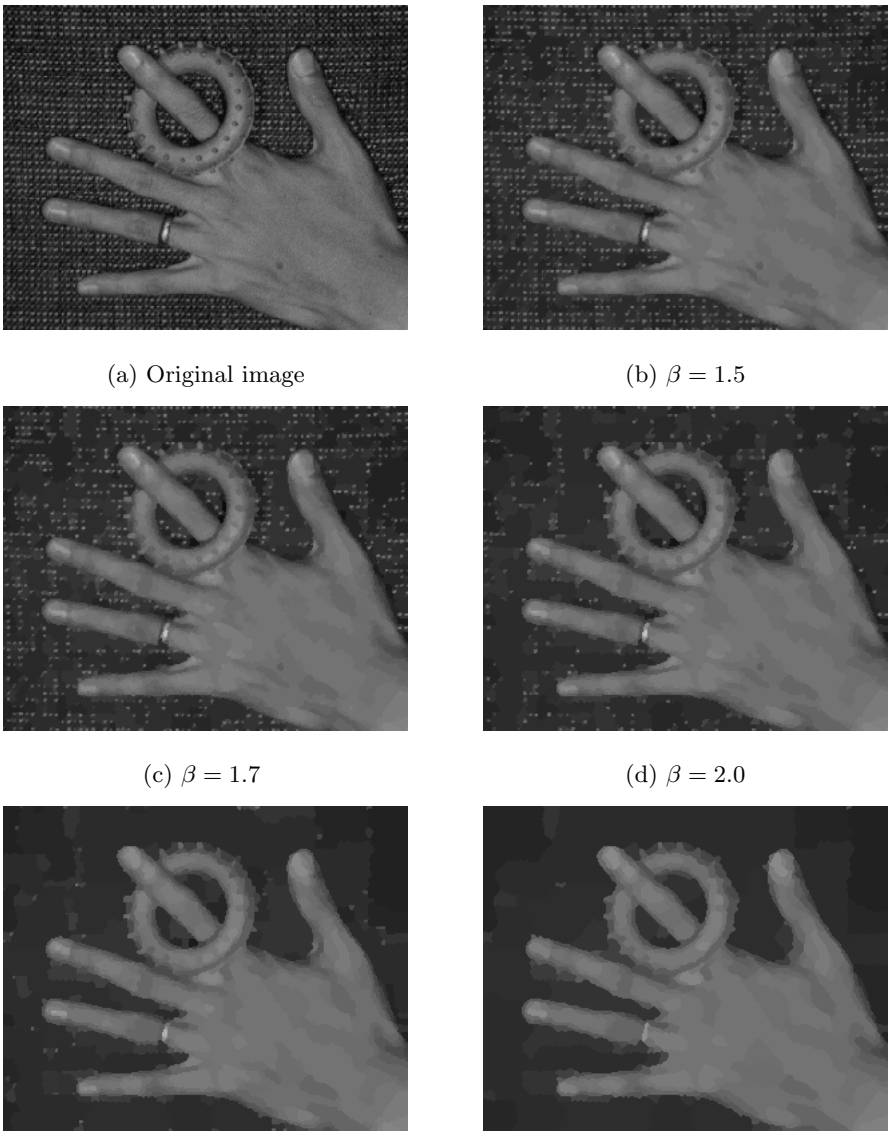


(d) Restoration using our method



(e) All level lines of (d)

Fig. 3. Restoration of a blocky image corrupted by a Gaussian noise. Results of TV minimization with L^2 fidelity for the gradient descent algorithm and our method. Only level lines multiples of 5 are displayed on (d).



(a) Original image

(b) $\beta = 1.5$

(c) $\beta = 1.7$

(d) $\beta = 2.0$

(d) $\beta = 2.5$

(e) $\beta = 3.0$

Fig. 4. Minimizers of TV with L^1 fidelity.