Exact Optimization of Discrete Constrained Total Variation Minimization Problems

Jérôme Darbon Marc Sigelle

October 18, 2004
N^o 2004C004
A paper based on this technical report has been accepted to IWCIA'2004 and published by Springer LNCS Series vol. 3322, pages 540–549

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 with 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.

Minimisation discrète et exacte de la variation totale sous contraintes

Jérôme Darbon Marc Sigelle

18 Octobre 2004
N^o 2004C004
Un papier utilisant les résultats de ce rapport a été accepté au congrès IWCIA'2004 et publié chez Springer LNCS Series 3322, pages 540–549

Résumé

Ce papier traite de la minimisation de la variation totale quand l'attache aux données est soit la norme L^2 ou soit la norme L^1. Nous proposons un algorithme qui fournit les solutions exactes ces deux problmes après une phase de discrtisation. Notre méthode repose sur la décomposition d'une image en ses lignes de niveaux. Les deux problèmes initiaux sont reformulés en termes de champs de Markov binaires associés chacun des ensembles de niveaux. Les solutions exactes de ces problèmes binaires sont exhibées grâce des algorithmes de coupures minimales. Nous prouvons que ces solutions présentent une relation de croissance ce qui permet de reconstruire la solution des deux problèmes initiaux.

1 Introduction

Image reconstruction and deconvolution methods are often based on the minimization of the constrained total variation [1, 2] of an image. These problems have minimizers in the space of functions of bounded variation [3] 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 (TV) of u is

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

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

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

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 $\sqrt{|\nabla u|^2 + \epsilon}$. 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 formulation of TV minimization using duality is presented in [4]. A fast algorithm which converges towards the solution can be derived from this formulation. In [5], 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 [6], 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 [7], 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 [8], Ishikawa presents an algorithm to find the exact solution for MRF 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:

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

where $\alpha \in \{1, 2\}$ and $\beta \geq 0$. The use of the L^1 fidelity has already been studied in [9, 10, 11]. 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 MRF framework. It 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 which belongs to the space of bounded variation, the Coarea formula [3] gives

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

for almost all λ and where $P(u^\lambda)$ is the perimeter of u^λ. In the discrete lattice version we define for each site s its grey level u_s and $u_s^\lambda = u^\lambda(s) = \mathbb{1}_{u_s \leq \lambda}$. We estimate the perimeter using pairs of neighboring pixels:

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

where $s \sim t$ denotes neighboring pixels and $R_{s,t}(u_s, v_s, \lambda) = w_{s,t}|u_s^\lambda - u_t^\lambda|$ ($w_{s,t}$ is some coefficient). For our experiments we use two different contour length estimators. The first one consists in considering only the four-connected neighborhood and setting $w_{s,t}$ to 1. The second one, as proposed in [12], 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 and L^2 through Level Sets

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

$$\begin{aligned}\sum_{s\in\Omega}|u_s - v_s| &= \sum_{u_s<v_s}|v_s - u_s| + \sum_{u_s>v_s}|u_s - v_s| = \sum_{s\in\Omega}\sum_{\lambda=0}^{N-1}(\mathbb{1}_{u_s\le\lambda<v_s} + \mathbb{1}_{v_s\le\lambda<u_s}) \\ &= \sum_{\lambda=0}^{N-1}\sum_{s\in\Omega}\mathbb{1}_{u_s\le\lambda}\,\mathbb{1}_{\lambda<v_s} + \mathbb{1}_{v_s\le\lambda}\,\mathbb{1}_{\lambda<u_s} = \sum_{\lambda=0}^{N-1}\sum_{s\in\Omega}u_s^\lambda\,(1-v_s^\lambda) + (1-u_s^\lambda)\,v_s^\lambda \\ &= \sum_{\lambda=0}^{N-1}\sum_{s\in\Omega}|u_s^\lambda - v_s^\lambda| = \sum_{\lambda=0}^{N-1}\sum_{s\in\Omega}D_1(u_s, v_s, \lambda) \qquad (1)\end{aligned}$$

where $D_1(x, y, \lambda) = |x^\lambda - y^\lambda|$, and where we used the property: $|a - b| = a + b - 2ab$ for binary variables a, b. 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.

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 previous disjoint sets and using the formula $\sum_{k=1}^{M}(2k-1) = M^2$:

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

Then 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:

$$\begin{aligned}\sum_{s\in\Omega}(u_s - v_s)^2 &= \sum_{u_s<v_s}\sum_{\lambda=u_s}^{v_s-1}(2(v_s-\lambda)-1) + \sum_{u_s>v_s}\sum_{\lambda=v_s}^{u_s-1}(2(\lambda-v_s)+1) \\ &= \sum_{s\in\Omega}\sum_{\lambda=0}^{N-1}\mathbb{1}_{u_s\le\lambda<v_s}\,(2(v_s-\lambda)-1) + \sum_{\lambda=0}^{N-1}\mathbb{1}_{v_s\le\lambda<u_s}\,(2(\lambda-v_s)+1) \\ &= \sum_{s\in\Omega}\sum_{\lambda=0}^{N-1}(\mathbb{1}_{u_s\le\lambda}\,\mathbb{1}_{\lambda<v_s} - \mathbb{1}_{v_s\le\lambda}\,\mathbb{1}_{\lambda<u_s})\,(2(v_s-\lambda)-1) \\ &= \sum_{\lambda=0}^{N-1}\sum_{s\in\Omega}(u_s^\lambda\,(1-v_s^\lambda) - (1-u_s^\lambda)\,v_s^\lambda)\,(2(v_s-\lambda)-1)\end{aligned}$$

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

where $D_2(x, y, \lambda) = (x^\lambda - y^\lambda)\,(2(y-\lambda)-1)$. This formulation shows that L^2 can be decomposed into level sets where their associated gray-levels are taken into account.

2.3 Independent Optimizations

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

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

where $E_\alpha^\lambda(u^\lambda,\beta) = \sum_{s\in\Omega} D_\alpha(u_s, v_s, \lambda) + \beta \sum_{s\sim t} w_{s,t}|u_s^\lambda - u_t^\lambda|$. Note that each term $E_\alpha^\lambda(u^\lambda,\beta)$ is a 2D MRF which only involves binary variables and pairwise interactions. Pairwise interactions only deal with the same gray-level component (λ) of two neighboring pixels (u_s and u_t). The data fidelity term can use different gray-level components of the observed image v, such as the L^2-norm case for instance. This is possible provided that the data fidelity energy can be linearly decomposed with respect to each component u_s^λ. The prior is an Ising model [13].

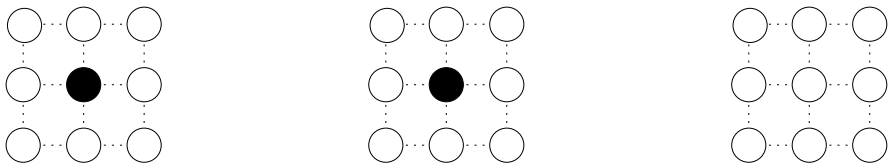
Now suppose that for each λ, we independently find the best binary configuration $\hat{u}^\lambda$ which minimizes the energy of the MRF. Clearly, the summation will be minimized. Thus we will find a minimizer for $E_\alpha(.,\beta)$ provided that the following property of monotony holds for binary minimizers:

$$\hat{u}^\lambda \leq \hat{u}^\mu \ \forall \lambda < \mu \quad . \qquad (3)$$

Indeed, if this property holds, then the minimizer $\hat{u}$ of $E_\alpha(.,\beta)$ is given [14] by $\hat{u}_s = \min\{\lambda, \hat{u}_s^\lambda = 1\}\ \forall s$. The monotone property is proved in the next section.

3 Reconstruction of the Solution

In this section, we prove the monotone property defined by (3). 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 the following Lemma will be useful.



(a) (b) (c)

Figure 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.

3.1 A Lemma based on coupled Markov Chains

Lemma *If the local conditional posterior energy at each site s can be written up to a constant, as:*

$$E_\alpha(u_s \mid \{u_t\},\ v_s) = \sum_{\lambda=0}^{N-1} \phi_s(\lambda)\ u_s^\lambda \tag{4}$$

where $\phi_s(\lambda)$ is a non-increasing function of λ, then one can exhibit a "coupled" stochastic algorithm minimizing each total posterior energy $E_\alpha^\lambda(u^\lambda,\beta)$ 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_α^k, there exists at least one solution $\hat{u}$ to the problem E_α^l such that $u^\star \leq \hat{u}\ \forall k \leq l$. The proof of the Lemma relies on coupled Markov chains [15].

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 (4) 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

$$P(u_s^\lambda = 1 \mid \{u_t^\lambda\},\ v_s^\lambda) = \frac{\exp -\phi_s(\lambda)}{1+\exp -\phi_s(\lambda)} = \frac{1}{1+\exp \phi_s(\lambda)}\ . \tag{5}$$

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 $u_s^\lambda = 0$.[1]
From the non-decreasing monotony of (5) 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 (5). Hence, this property also holds for a series of N coupled Simulated Annealing algorithms [16] where a *single* temperature T boils down to 0 (either after each visited site s or at the beginning of each tour [13] .) $\square$
Several points should be emphasized here:

- The coupled monotony-preserving Gibbs samplers described in [15] relate to the *same* MRF but for various initial conditions, while here, our N coupled Gibbs samplers relate to N *different* posterior MRF's (one for each level λ).

- 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.

3.2 The L^1 and L^2 cases

Let us show that both L^1 regularization and attachment to data energies feature property (4); so will do their sum and thus the total posterior energy. Using previous property for binary variables a, b: $|a-b| = a+b-2ab$, this yields:

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

where $C = \sum_{\lambda=0}^{N-1} \sum_{t \sim s} u_t^\lambda$ is a "constant" since it only depends on the $\{u_t^\lambda\}$.
Thus $\phi_s(\lambda) = \sum_{t \sim s} (1 - 2u_t^\lambda)$, which is by essence a non-increasing function of λ.

Similarly starting from (1) for the L^1 attachment to data term:

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

[1] 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.

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

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

The approach for the L^2 relies on the same method. From (2) one can write:

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

where $\phi_s(\lambda) = 2(v_s - \lambda) - 1$ clearly fulfills our requirement.

Thus in both cases, $\phi_s(\lambda)$ is a non-increasing function, so that TV regularization with either L^1 and L^2-fidelity both follow the conditions of our Lemma. 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 it has no stopping criteria. We propose an algorithm in the next section.

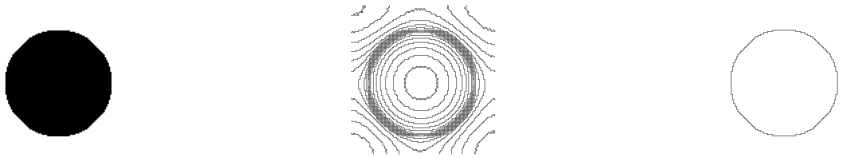
4 Computations and Experiments

In this section, we describe our algorithm and present some experiments.

4.1 Minimum-cut Based Minimization

Greig et al. [17] were the first ones to propose an exact optimization for binary MRF. It is based on constructing a graph such that its minimum cut (MC) gives an optimal labelling. Since this seminal work, other graph constructions were proposed to solve some non-binary problems exactly [8, 18]. In [19], the authors propose a necessary condition for binary functions to be minimized via MCs along with a graph construction. Our Ising model fulfills the condition.

For each level we construct the graph as proposed in [19] and compute a MC. 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. To reconstruct the solution, one flips every pixel where this property is violated. This flipping process also gives an optimal labelling since the energy does not change.



(a) (b) (c)

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

To compute the MC, we used the algorithm described in [20]. For our binary problems, this algorithm gives near-linear performance with respect to the number of pixels $|\Omega|$. Since we compute N cuts, the complexity of our algorithm is near-linear both with respect to N and $|\Omega|$. Time results (on a 1.6GHz Pentium IV) for our method are presented in table 1 for L^1 fidelity. This is in contrast with the near-quadratic behavior of [8] with respect to N.

4.2 Experiments

For our experiments, we always use the 8-connectivity. In [21] 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. Figure 2 depicts the level-lines of the solutions for our algorithm and the gradient descent algorithm. For the latter, we approximate $TV(u)$ by $\sqrt{|\nabla u|^2 + \epsilon}$ with $\epsilon = 1$. Note how many level lines are created by the gradient descent algorithm.

TV minimization is 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 [22] which states that a square cannot arise as a solution. Results of the regularization using L^1-fidelity are depicted in figure 4. The higher the coefficient β, the more fine structures are removed while the contrast remains 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. We will show that the condition stated by our Lemma is equivalent to the fact that each local conditional posterior energy is a convex function. 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] Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D. **60** (1992) 259–268

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

[3] Evans, L., Gariepy, R.: Measure Theory and Fine Properties of Functions. CRC Press (1992)

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

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

[6] Pollak, I., Willsky, A., Huang, Y.: Nonlinear evolution equations as fast and exact solvers of estimation problems. to appear in IEEE Signal Processing (2004)

[7] Amini, A., Weymouth, T., Jain, R.: Using dynamic programming for solving variational problems in vision. IEEE PAMI **12** (1990) 855–867

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

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

[10] Chan, T., Esedoğlu, S.: Aspect of total variation regularized l^1 function approximation. Technical Report 7, UCLA (2004)

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

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

[13] Winkler, G.: Image Analysis, Random Fields and Dynamic Monte Carlo Methods. Applications of mathematics. Springer-Verlag (2003)

[14] Guichard, F., Morel, J.: Mathematical morphology "almost everywhere". In: Proceedings of ISMM, Csiro Publishing (2002) 293–303

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

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

[17] Greig, D., Porteous, B., Seheult, A.: Exact maximum a posteriori estimation for binary images. Journal of the Royal Statistics Society **51** (1989) 271–279

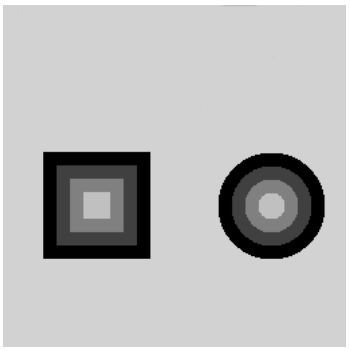
[18] Roy, S.: Stereo without epipolar lines: A maximum-flow formulation. International Journal of Computer Vision **34** (1999) 147–162

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

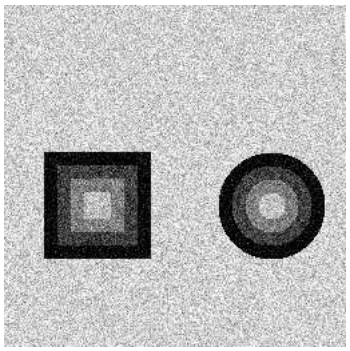
[20] Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE PAMI **26** (2004) 1124–1137

[21] Strong, D., Chan, T.: Edge preserving and scale-dependent properties of total variation regularization. Inverse Problem **19** (2003) 165–187

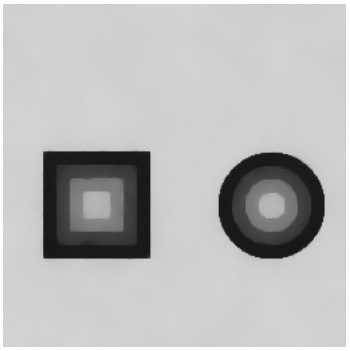
[22] Meyer, Y.: Oscillating patterns in image processing and nonlinear evolution equations. University Lecture Series **22** (2001)



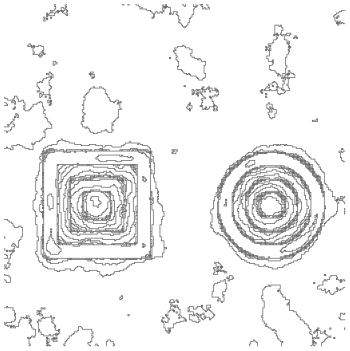
(a) Original image



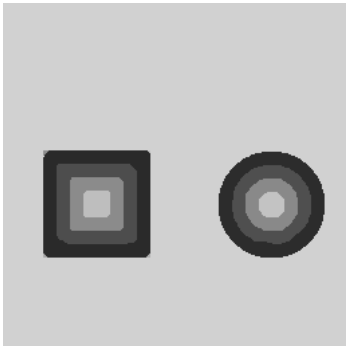
(b) Noisy image ($\sigma = 30$)



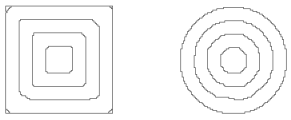
(c) Gradient descent restoration



(d) Some level lines of (c)



(d) Restoration using our method



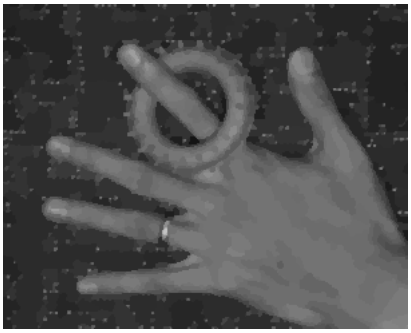
(e) All level lines of (d)

Figure 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) $\beta = 3.0$

Figure 4: Minimizers of TV with L^1 fidelity.