FINDING THE GLOBAL MINIMUM FOR BINARY IMAGE RESTORATION

Tony F. Chan†*, Selim Esedoḡlu*† *and Mila Nikolova*⋆

† Mathematics Department, UCLA. Los Angeles, CA 90095.
⋆ Centre de Mathematiques et de Leurs Applications, ENS de Cachan, France.

ABSTRACT

Restoring binary images is a problem which arises in various application fields. In our paper, this problem is considered in a variational framework: the sought-after solution minimizes an energy. Energies defined over the set of the binary images are inevitably nonconvex and there are no general methods to calculate the global minimum, while local minimziers are very often of limited interest. In this paper we define the restored image as the global minimizer of the total-variation (TV) energy functional constrained to the collection of all binary-valued images. We solve this constrained non-convex optimization problem by deriving another functional which is convex and whose (unconstrained) minimum is proven to be reached for the global minimizer of the binary constrained TV functional. Practical issues are discussed and a numerical example is provided.

1. INTRODUCTION

In various applications we are given a binary-valued function $f(x)$: $\mathbf{R}^N \to \{0,1\}$, $N \geq 2$, which is known to be the corrupted version of another binary-valued function $u(x)$ that needs to be estimated. We can evoke text denoising and document processing, two-phase image segmentation, shape restoration, channel-noise cancellation in communications, fairing of surfaces in computer graphics and others. This problem can be stated either as denoising or as segmentation. Since both u and f are binary, they can be represented as the characteristic function of a shape. The corruption incurred by f is thus in the geometry of the shape: Its boundary might be very rough, and the user might be interested in smoothing out its boundary, and perhaps removing small, unnecessary connected components of the shape. This task is a common first step in many object detection and recognition algorithms. Variational and partial differential equations based approaches to denoising and segmentation have had great success, essentially because these models are well suited to impose geometric regularity on the solutions sought. Among the best known and most influential examples are the Rudin-Osher-Fatemi (ROF) total variation based image denoising model, and the Mumford-Shah image segmentation model. These models are easily adapted to binary images. However, a common difficulty that arises is the presence of spurious local minima that are not global minima. The reason is that the constraint set—the family of all characteristic functions of subsets of $\mathbf{R}^N$—is a non-convex collection. This is a much more serious drawback than non-uniqueness of global minimizers because local minima of segmentation and denoising models often have completely wrong levels of detail and scale: whereas global minimizers of a given model are usually all reasonable solutions, the local minima tend to be blatantly false. Many solution techniques for variational models are based on gradient descent, and are therefore prone to getting stuck in such local minima. This makes initial guess for gradient descent based algorithms sometimes critically important for obtaining satisfactory results.

In this paper we propose a method that guarantees to reach the global minimum of the ROF model restricted to the set of binary images. Our approach is to consider the *unconstrained* minimization of a *convex* functional whose minimum is reached for the constrained global minimizer of the ROF model. The theory in this work relies on the results established in [1]. A similar idea in the simpler context of images on a finite grid was used in [2] in order to find quasi-binary solutions by minimizing a convex functional.

2. BINARY IMAGE RESTORATION USING CONSTRAINED MINIMIZATION

Let $f(x) : \mathbf{R}^N \to [0,1]$ denote the given (grayscale) possibly corrupted (noisy) image. Since [3], an usual approach for image restoration is to minimize an energy of the form

$$E(u) = \int_{\mathbf{R}^N} \varphi(|\nabla u|) + \lambda \int_{\mathbf{R}^N} \Big(u(x) - f(x)\Big)^2 dx \quad (1)$$

where $\lambda > 0$ is a parameter to be chosen by the user, or estimated if the level of noise is known and $\varphi : \mathbf{R}_+ \to \mathbf{R}$ is a function [4, 5, 6, 7]. One of the most popular examples is $\varphi(t) = t$ which corresponds to the Rudin-Osher-Fatemi (ROF) total variation based image denoising model [4]. In the *binary* image denoising case, $f(x)$ can be expressed as

$$f(x) = \mathbf{1}_\Omega(x)$$

where Ω is a bounded subset of $\mathbf{R}^N$ whose boundary $\partial\Omega$ can be very rough because of the noise. A natural way to solve such a problem is to constraint the unknown u in (1) to be binary, i.e. to have the form $u(x) = \mathbf{1}_\Sigma(x)$. This possibility was considered in [8]. For the ROF model one then obtains the following constrained optimization problem:

$$\min_{\substack{\Sigma \subset \mathbf{R}^N \\ u(x) = \mathbf{1}_\Sigma(x)}} \underbrace{\int_{\mathbf{R}^N} |\nabla u| + \lambda \int_{\mathbf{R}^N} \Big(u(x) - \mathbf{1}_\Omega(x)\Big)^2 dx}_{E_2(u)} \quad (2)$$

Problem (2) is non-convex because the minimization is over a non-convex set of functions. Notice that (2) in equivalent to the following geometry problem:

$$\min_{\Sigma \subset \mathbf{R}^N} \mathrm{Per}(\Sigma) + \lambda |\Sigma \bigtriangleup \Omega| \quad (3)$$

where $\mathrm{Per}(\cdot)$ denotes the perimeter, $|\cdot|$ is the N-dimensional Lebesgue measure, and $S_1 \bigtriangleup S_2$ denotes the symmetric difference between

0-7803-9134-9/05/$20.00 ©2005 IEEE

the two sets S_1 and S_2. The unknown set Σ in (3) can be described by its boundary $\partial\Sigma$. So a common approach of solving (3) has been to use some *curve evolution* process, sometimes referred to as *active contours*, where $\partial\Sigma$ is updated iteratively, usually according to gradient flow for the energy involved.

Numerically, there are several ways of representing $\partial\Sigma$. Explicit curve representations as in Kass, Witkin, Terzopoulos [9] are not appropriate since such methods do not allow changes in curve topology (and have a number of other drawbacks). Actually, the most successful algorithms are those based on either the level set method of Osher and Sethian [10], or on the variational approximation approach known as Gamma convergence theory [11]. In the level set formulation, $\partial\Sigma$ is represented as the 0-level set of a (Lipschitz) function $\phi : \mathbf{R}^N \to \mathbf{R}$: $\Sigma = \left\{x \in \mathbf{R}^N : \phi(x) > 0\right\}$ so that $\partial\Sigma = \{x \in \mathbf{R}^N : \phi(x) = 0\}$. The functional E_2 in (2) can then be expressed as follows:

$$\int_{\mathbf{R}^N} |\nabla H(\phi(x))|\, dx + \lambda \int_{\mathbf{R}^N} \Big(H(\phi(x)) - \mathbf{1}_\Omega(x)\Big)^2 dx \quad (4)$$

where $H : \mathbf{R} \to \mathbf{R}$ is the Heaviside function, $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x \geq 0$. In practice, one takes a smooth (or at least Lipschitz) approximation H_ε to H, such that $H_\varepsilon(x) \to H(x)$ as $\varepsilon \to 0$. The Euler-Lagrange equation for (4) leads to the following gradient flow:

$$\phi_t(x,t) = H'_\varepsilon(\phi)\left\{\mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + 2\lambda\Big(\mathbf{1}_\Omega(x) - H_\varepsilon(\phi)\Big)\right\}. \quad (5)$$

When (5) is simulated using reinitialization for the level set function $\phi(x)$ and a compactly supported approximation $H_\varepsilon(x)$ to $H(x)$, it is observed to define a continuous evolution (with respect to, say, the L^1-norm) for the unknown function $u(x) = \mathbf{1}_\Sigma(x)$ and decreases the objective energy (2) through binary images.

In the Gamma-convergence approach, E_2 is replaced by a sequence of approximate energies E_ε of the form

$$\begin{aligned} E_\varepsilon(u) &= \int_{\mathbf{R}^N} \varepsilon|\nabla u|^2 + \frac{1}{\varepsilon}W(u) \\ &+ \lambda\Big\{u^2\big(c_1 - f\big)^2 + (1-u)^2\big(c_2 - f\big)^2\Big\}\, dx \end{aligned}$$

where $E_\varepsilon \to E_2$ as $\varepsilon \to 0$. Here, $W(\xi)$ is a double-well potential with equidepth wells at 0 and 1; e.g., a simple choice is $W(\xi) = \xi^2(1-\xi)^2$. The term $\frac{1}{\varepsilon}W(u)$ is a penalty that forces the function u to be approximately 0 or 1 on most of $\mathbf{R}^N$. The term $\varepsilon|\nabla u|^2$, on the other hand, puts a penalty on the transitions of u between 0 and 1. Taken together, these terms constraint u to be a characteristic function, and approximate its total variation. For rigorous details, see [12].

However, these techniques are prone to get stuck in spurious local minima, thus leading to images with wrong level of detail. This fact is familiar to researchers working with these techniques and is confirmed by the example below.

Example: Let $f(x) = \mathbf{1}_\Omega(x)$ where $\Omega = B_R(0) \subset \mathbf{R}^2$ is a ball of radius R centered at the origin. Implementing the gradient descent algorithm defined by (5) requires to choose an initial guess for the set Σ that is represented by $\phi(x)$. A common choice being to take the observed image, we initially set $\Sigma = B_R(0)$. The evolution defined by (5) will maintain radial symmetry of $\phi(x)$. That is, at any given time $t \geq 0$, the set represented by $\phi(x)$ (i.e. the candidate for minimization) is of the form $\Big\{x \in \mathbf{R}^2 : \phi(x) > 0\Big\} = B_r(0)$ for some radius $r \geq 0$. We can write the energy of $u(x) = \mathbf{1}_{B_r(0)}(x)$ in terms of r, as follows: $E(r) := E_2(\mathbf{1}_{B_r(0)}(x)) = 2\pi r + \lambda\pi|R^2 - r^2|$ A simple calculation shows that if $0 < \lambda < \frac{2}{R}$, then the minimum of this function is at $r = 0$. Hence, the denoising model prefers to remove disks of radius smaller than the critical value $\frac{2}{R}$. If in addition $R > \frac{1}{\lambda}$, it is easy to find that $E(r)$ has a local maximum at $r_{max}(\lambda) = \frac{1}{\lambda}$. (see Figure 1). Thus the energy minimization procedure described by (5) cannot shrink disks of radius $R \in (\frac{1}{\lambda}, \frac{2}{\lambda})$ to a point, even though the global minimum of the energy for an original image given by such a disk is at $u(x) \equiv 0$.

We can easily say a bit more: There is $\delta > 0$ such that if $\Sigma \subset \mathbf{R}^N$ satisfies $|\Sigma \triangle B_R(0)| < \delta$ then $E_2(\mathbf{1}_\Sigma(x)) > E_2(\mathbf{1}_{B_R(0)}(x))$. In words, all binary images nearby, but not identical with, the observed image $\mathbf{1}_{B_R(0)}(x)$ have strictly higher energy.

To summarize: If $f(x) = \mathbf{1}_{B_R(0)}(x)$ with $R \in (\frac{1}{\lambda}, \frac{2}{\lambda})$, and if the initial guess for the continuous curve evolution based minimization procedure (5) is $f(x)$, then the procedure gets stuck in the local minimizer $u(x) = f(x)$ while the unique global minimizer is $u(x) \equiv 0$.

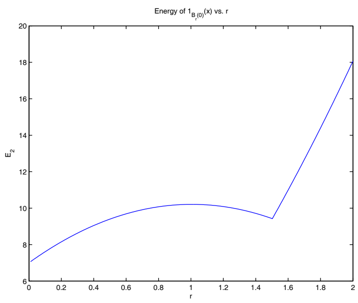

Fig. 1. Energy (2) of $u(x) = \mathbf{1}_{B_r(0)}(x)$ as a function of $r \in [0, 2]$ when the observed image is given by $f(x) = \mathbf{1}_{B_R(0)}(x)$. Here, $R = \frac{3}{2}$ and the parameter λ was chosen to be $\lambda = 1$. There is clearly a local minimum, corresponding to $r = R = \frac{3}{2}$.

3. FINDING THE GLOBAL MINIMUM USING ANOTHER (CONVEX) ENERGY

The crux of our approach is to consider minimization of the following convex energy, defined for any given observed image $f(x) \in L^1(\mathbf{R}^N)$ and $\lambda \geq 0$:

$$E_1(u) := \int_{\mathbf{R}^N} |\nabla u| + \lambda \int_{\mathbf{R}^N} |u(x) - f(x)|\, dx \quad (6)$$

This energy differs from the standard ROF model in the fidelity term: The L^2-norm square of the original model is replaced by the L^1-norm as a measure of fidelity. It was previously introduced and studied in signal and image processing applications in [13, 14, 15, 16, 17, 1]. The energy E_1 has many interesting properties and uses [1]. It is easy to show that the set of the minimizers of E_1 is non-empty, closed and convex.

The relevance of (6) for our purposes comes from the fact that E_1 is *convex*, hence its minimum can practically be reached, and from the equivalence theorem stated below.

Theorem 1 (Equivalence) *Let $f = \mathbf{1}_\Omega$ where $\Omega \subset \mathbf{R}^N$ is a bounded domain. Then we have the following:*

(i) If $v = \mathbf{1}_\Sigma$ is a (global) solution to (2), then E_1 reaches its minimum at v.

(ii) If E_1 reaches its minimum at w, then for almost every $\mu \in (0,1)$ the function $v = \mathbf{1}_\Sigma$ where $\Sigma = \{x \in \mathbf{R}^N : w(x) > \mu\}$ is a (global) solution to (2).

These statements come from the obvious fact that the energies in (2) and (6) agree on binary images, i.e. $E_2(\mathbf{1}_\Sigma) = E_1(\mathbf{1}_\Sigma)$ for any bounded $\Sigma \subset \mathbf{R}^N$, and from Theorem 2 which is taken from [1]:

Theorem 2 *If $f = \mathbf{1}_\Omega$ where $\Omega \subset \mathbf{R}^N$ in bounded, then there is $\Sigma \subset \mathbf{R}^N$ (possibly $\Sigma \neq \Omega$) such that E_1 reaches its minimum at $v = \mathbf{1}_\Sigma$.*

More precisely, if w is any minimizer of E_1, then for almost every $\mu \in [0,1]$ the function $v = \mathbf{1}_\Sigma$ where $\Sigma = \{x : w(x) > \mu\}$ is also a minimizer of E_1.

Its proof is based on the following proposition, established in [1], that expresses energy E_1 in (6) in terms of the level sets of u and f.

Proposition 1 *The energy E_1 can be rewritten as follows:*

$$E_1(u) = \int_{-\infty}^{\infty} Per\left(\{x : u(x) > \mu\}\right)$$
$$+\lambda \Big| \{x : u(x) > \mu\} \triangle \{x : f(x) > \mu\} \Big| \, d\mu$$

Since E_1 is only non-strictly convex, its (global) minimizers are not unique in general. Nevertheless, all its minimizers are global. By the statements above, E_1 necessarily has a binary minimizer: if there is a non-binary minimizer, it is then nonstrict and it is connected with another minimizer which is binary.

Algorithm To find a solution (i.e. a global minimizer) $v(x)$ of the non-convex variational problem (2), it is sufficient to carry out the following three steps:

1. Find any minimizer $w(x)$ of the **convex** energy (6).
2. Determine $\Sigma = \{x \in \mathbf{R}^N : w(x) > \mu\}$ for some $\mu \in (0,1)$.
3. Set $w(x) = \mathbf{1}_\Sigma(x)$: then v is a global minimizer of (2) for almost every choice of μ.

The most involved step in the solution procedure above is finding a minimizer of (6). One can approach this problem in many ways; for instance, one possibility is to simply carry out subgradient descent. Further details can be found in a recent report of the authors [18].

On the Premises of our Work

Our analysis is inherently related to the following result of Strang [19, 20]: for f a given function, the solutions of the constrained minimization problem

$$\inf_{\{u : \int f u \, dx = 1\}} \int |\nabla u| \tag{7}$$

are characteristic functions of sets. The main idea involved is to express both the functional to minimize and the constraint in terms of the level sets of u and f. The coarea formula of Fleming and Rishel [21] is the primary tool. In our work, we apply the idea of expressing the functionals in terms of level sets. However, our problem is "opposite" to that of [20, 19] in the sense that we are looking for functionals whose minimizers are characteristic functions.

4. NUMERICAL EXAMPLE

We show a sample computation on a synthetic image. The image of Figure 2 represents the given binary image $f(x)$, which is a simple geometric shape covered with random (binary) noise. The initial guess was an image composed of all 1's (an all white image). In the computation, the parameter λ was chosen to be quite moderate, so that in particular the small circular holes in the shape should be removed while the larger one should be kept. The result of the minimization is shown in Figure 3; in this case the minimizer is automatically very close to being binary, and hence the thresholding step of the algorithm in Corollary 3 is almost unnecessary.

Figure 4 shows the histograms of intermediate steps during the gradient descent based minimization. As can be seen, the intermediate steps themselves are very far from being binary. The histogram in the lower right hand corner belongs to the final result shown in Figure 3. Thus the gradient flow goes through non-binary images, but in the end reaches another binary one. Although this is not implied by Theorem 2, it seems to hold in practice.

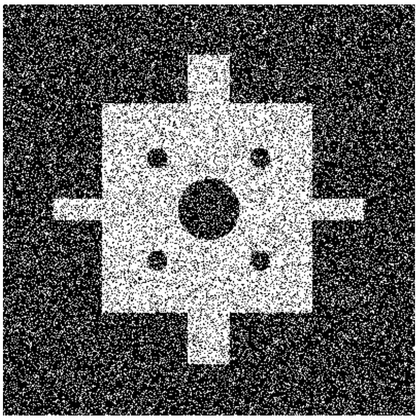

Fig. 2. Original binary image.

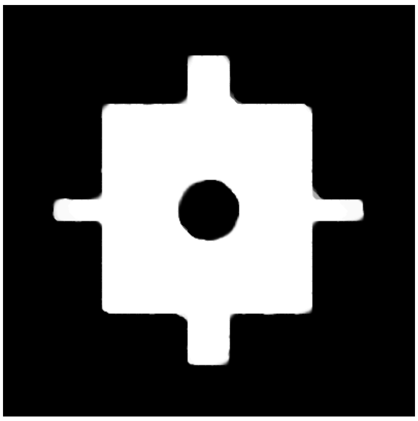

Fig. 3. Final result found (no need to threshold in this case).

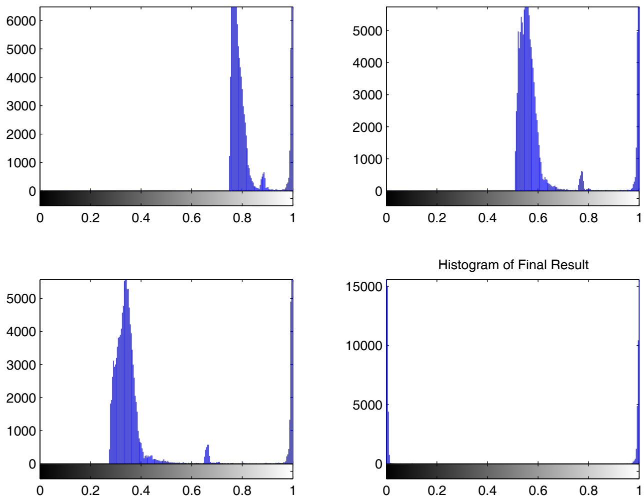

Fig. 4. Histograms for intermediate images as the gradient descent proceeds. As can be seen, the intermediate images themselves are not binary; however, by the time the evolution reaches steady state, we are back to a binary image.

5. CONCLUSIONS AND PERSPECTIVES

in this work we provide a convergent method how to solve the non-convex problem of the finding of a minimizer of the regularized total-variation functional constrained to the collection of all binary-valued images. The key point is to propose a *convex* functional that is minimized for the sought-after solution. These results can be extended to piecewise constant Mumford-Shah segmentation energy [22], which requires extension to given images f that are not binary. However, we will not dwell on this further here.

6. REFERENCES

[1] Tony Chan and Selim Esedoglu, "Aspects of total variation regularized l^1 function approximation," Tech. Rep., CAM Report 04-07, 2004, to appear in SIAM Journal on Applied Mathematics.

[2] Mila Nikolova, "Estimation of binary images using convex criteria," in *Proceedings of the IEEE International Conference on Image Processing*, 1998, vol. 2, pp. 108–112.

[3] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Winston, Washington DC, 1977.

[4] L. Rudin, S. Osher, and C. Fatemi, "Nonlinear total variation based noise removal algorithm," *Physica*, vol. 60 D, pp. 259–268, 1992.

[5] Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Transactions on Image Processing*, vol. 6, no. 2, pp. 298–311, Feb. 1997.

[6] G. Aubert and L. Vese, "A variational method in image recovery," *SIAM Journal on Numerical Analysis*, vol. 34, no. 5, pp. 1948–1979, 1997.

[7] Gilles Aubert and Pierre Kornprobst, *Mathematical problems in images processing*, Springer-Verlag, Berlin, 2002.

[8] Luminita Vese and Stanley Osher, "The level set method links active contours, mumford-shah segmentation, and total variation restoration," Tech. Rep., UCLA CAM Report (02-05), 2002.

[9] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, , no. 1, pp. 321–331, 1987.

[10] S. Osher and J. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.

[11] G. Dal Maso, *An introduction to* Γ*-convergence*, Progress in Nonlinear Differential Equations and their Applications. Birkhauser Boston Inc., Boston, MA, 1993.

[12] L. Modica and S. Mortola, "Un esempio di gamma-convergenza," *Boll. Un. Mat. Ital. B (5)*, vol. 14, 1 1977.

[13] S. Alliney, "Digital filters as absolute norm regularizers," *IEEE Transactions on Signal Processing*, vol. SP-40, no. 6, pp. 1548–1562, 1992.

[14] S. Alliney, "Recursive median filters of increasing order: a variational approach," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1346–1354, 1996.

[15] S. Alliney, "A property of the minimum vectors of a regularizing functional defined by means of absolute norm," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 913–917, Apr. 1997.

[16] Mila Nikolova, "A variational approach to remove outliers and impulse noise," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, Jan. -Mar. 2004.

[17] Mila Nikolova, "Minimizers of cost-functions involving non-smooth data-fidelity terms. Application to the processing of outliers," *SIAM Journal on Numerical Analysis*, vol. 40, no. 3, pp. 965–994, 2002.

[18] Tony Chan, Selim Esedoglu, and Mila Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," Tech. Rep., CAM Report 04-54, 2004.

[19] Gilbert Strang, "Maximal flow through a domain," *Mathematical Pogramming*, vol. 26, no. 2, pp. 123–143, 1983.

[20] Gilbert Strang, L^1 *and* L^∞ *approximation of vector fields in the plane*, pp. 273–288, North-Holland Math. Stud. 81, North Holland, Amsterdam, 1983.

[21] W. Fleming and R. Rishel, "An integral formula for total gradient variation," vol. 11, 1960.

[22] Tony Chan and Luminita Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.