An Algorithm for Minimizing the Mumford-Shah Functional*

Thomas Pock
Graz University of Technology, Austria

Daniel Cremers
University of Bonn, Germany

Horst Bischof
Graz University of Technology, Austria

Antonin Chambolle
Ecole Polytechnique, Paris, France

Abstract

In this work we revisit the Mumford-Shah functional, one of the most studied variational approaches to image segmentation. The contribution of this paper is to propose an algorithm which allows to minimize a convex relaxation of the Mumford-Shah functional obtained by functional lifting. The algorithm is an efficient primal-dual projection algorithm for which we prove convergence. In contrast to existing algorithms for minimizing the full Mumford-Shah this is the first one which is based on a convex relaxation. As a consequence the computed solutions are independent of the initialization. Experimental results confirm that the proposed algorithm determines smooth approximations while preserving discontinuities of the underlying signal.

1. Introduction

In 1989 Mumford and Shah [16] suggested to minimize the functional

$$E(u) = \lambda \int_{\Omega} (f - u)^2 dx + \int_{\Omega \setminus S_u} |\nabla u|^2 dx + \nu \mathcal{H}^1(S_u), \quad (1)$$

in order to approximate an input image $f : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ in terms of a piecewise smooth function $u : \Omega \rightarrow \mathbb{R}$. The functional contains a data fidelity term and two regularity terms imposing smoothness of u in areas separated by the discontinuity set S_u and regularity of S_u in terms of its one-dimensional Hausdorff measure $\mathcal{H}^1(S_u)$. Related approaches were proposed in a spatially discrete setting by Geman and Geman [12] and by Blake and Zisserman [4]. To date the paper of Mumford and Shah has attracted more than 1600 citations. In practice, one of the major challenges is to

---

develop efficient algorithms to compute high quality minimizers of this functional. In the following we will briefly review a number of the most popular algorithms and discuss their shortcomings.

(a) Input image   (b) Approximation

Figure 1. Approximation of a natural image using the proposed algorithm for minimizing the piecewise smooth Mumford-Shah functional. Our method leads to high quality solutions being independent from the initialization.

1.1. Related Work

Minimizing the Mumford-Shah functional in its original setting turns out to be a challenging problem due to the non-regularity of the edge term. It is therefore not astonishing that there have been several lines of research dealing with the question of how to find good approximations of the Mumford-Shah functional.

One of the earliest attempts are based on so-called continuation methods, such as simulated annealing [12] or the graduated non convexity (GNC) procedure [4]. The idea is to minimize the original energy by gradually decreasing a continuation parameter. However, the performance of these methods largely depend on the dynamics of the continuation parameter and therefore tend to get stuck in bad local minima.

A quite popular class of algorithms is based on the level set method [20, 19] or splines [10]. While these meth-

ods work well for certain image segmentation tasks, they are limited in several ways. Firstly, they do not allow the formation of open boundaries (also known as crack tips) since level sets always being closed regions, Secondly, the curves are propagated only locally and therefore can easily get stuck in local minima (e.g. they do not allow to detect interior boundaries). Lastly, in the case of multiple regions regions, the cost functional overcounts the boundary length in those regions where multiple zero level sets meet. Alternatively the level set framework can be replaced by graph cut algorithms [14] or convex relaxation approaches [9, 7]. While these methods allow to find good solutions with respect to the curve evolution (for two regions they can even find the globally optimal solution), these methods still do not minimize the full Mumford-Shah energy, as we do so in this work. Similar to level sets, graph based approaches also do not allow to represent open boundaries.

A special case of the full Mumford-Shah functional - the piecewise constant Mumford-Shah functional - is obtained by setting the weight of the smoothness term in (1) to $+\infty$. This results in a piecewise constant approximation of the input image f. Very recently Pock et al. proposed in [17] a convex relaxation approach which allows to compute high quality solutions of the piecewise constant Mumford-Shah functional.

In [3], Ambrosio and Tortorelli proposed to approximate the original Mumford-Shah energy (1) by a sequence of simpler elliptic variational problems. They proposed to replace the edge set S_u by means of a 2D function z and designed the so-called phase field energy

$$L_{z,\varepsilon} = \int_\Omega \varepsilon |\nabla z|^2 \mathbf{dx} + \int_\Omega \frac{(1-z)^2}{4\varepsilon} \, \mathbf{dx} \, . \qquad (2)$$

The remarkable property associated with this formulation is that as $\varepsilon \to 0$, $L_{z,\varepsilon}$ approaches the length of S_u, i.e. $\mathcal{H}^1(S_u)$. However, this approximation works well only if the scale of ε is in the order of the size of the pixel grid. On the other hand, the Ambrosio-Tortorelli approximation can handle open boundaries and therefore being more in the spirit of the original formulation.

In [13, 8] the authors presented a non-local approximation of the Mumford-Shah functional which is inspired by the original discrete model of Blake and Zisserman [4]. The major advantage of this formulation is, that the explicit computation of the jump set S_u is avoided by using a family of continuous and non-decreasing functions $f : [0, +\infty) \to [0, +\infty)$ satisfying

$$\lim_{t \to 0^+} \frac{f(t)}{t} = 1 \, , \quad \lim_{t \to +\infty} f(t) = 1 \, .$$

While Chambolle uses functions of the form $f(t) = \arctan(t)$, it was later shown that $f(t) = log(1 + t)$ is a better choice, since it is less sensitive to local minima and

requires a smaller number of iterations to converge [15]. Let us finally mention that this types of approximations work well in practice but minimize only an anisotropic variant of the Mumford-Shah functional.

1.2. Contribution

In this paper we propose a novel algorithm to compute high quality solutions of the piecewise smooth Mumford-Shah functional. Our approach is based on a convex representation due to Bouchitte, Alberti and Dal Maso. As the central contribution of our work, we propose a novel fast and convergent primal-dual algorithm to compute the solution of the convex representation of the Mumford-Shah functional. Basically, computing the solution of the Mumford-Shah functional amounts for a saddle point search in higher dimensions. We compare the performance of our algorithm to an old algorithm proposed due to Popov and a recently proposed primal-dual algorithm for which convergence cannot be proven. It turns out that our new primal-dual algorithm clearly outperforms both methods. In experimental results on synthetic and real images we compare our algorithm to a non convex method of Ambrosio and Tortorelli. We demonstrate that our algorithm leads to more precise solutions while being independent of the initialization.

2. Convex Relaxation for the Mumford-Shah Functional

Let $\Omega \subset \mathbb{R}^2$ denote the image plane and let $u \in SBV(\Omega)$[1] Denote the upper level sets of u by the characteristic function $\mathbf{1}_u : \Omega \times \mathbb{R} \to \{0, 1\}$ of the subgraph of u:

$$\mathbf{1}_u(x, t) = \begin{cases} 1, & \text{if } t < u(x), \\ 0, & \text{else.} \end{cases} \qquad (3)$$

In a series of papers Bouchitte, Alberti, Dal Maso [5, 1] proposed convex relaxations for the Mumford-Shah functional. The main result of their work is summarized by the following theorem.

Theorem 1. *For a function $u \in SBV(\Omega)$ the Mumford-Shah functional can be written as*

$$E(u) = \sup_{\varphi \in K} \int_{\Omega \times \mathbb{R}} \varphi D\mathbf{1}_u, \qquad (4)$$

---

[1] $SBV(\Omega)$ denotes the special functions of bounded variation [2], i.e. functions u of bounded variation for which the derivative Du is the sum of an absolutely continuous part $\nabla u \cdot dx$ and a discontinuous singular part S_u – see Figure 2.
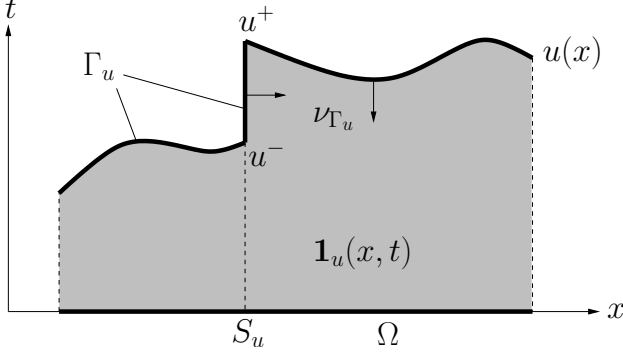
Figure 2. Schematic plot of a function $u \in SBV(\Omega)$ showing the continuous and discontinuous part. The convex relaxation of the Mumford-Shah functional is based on implicitly representing functions u by the characteristic function $\mathbf{1}_u$ of its subgraph which is 1 only in the gray shaded area under the curve. A convex approximation of the Mumford-Shah functional (1) is obtained by maximizing the flux through the interface Γ_u over all vector fields in an appropriately chosen convex set.

with a convex set

$$K = \Big\{ \varphi \in C_0(\Omega \times \mathbb{R}; \mathbb{R}^2) :$$

$$\varphi^t(x,t) \geq \frac{\varphi^x(x,t)^2}{4} - \lambda(t - f(x))^2, \quad (5)$$

$$\left| \int_{t_1}^{t_2} \varphi^x(x,s)ds \right| \leq \nu \Big\},$$

where the inequalities in the definition of K hold for all $x \in \Omega$ and for all $t, t_1, t_2 \in \mathbb{R}$.

In particular, if for a given u the supremum in (4) is attained by a *divergence-free* vector field $\bar{\varphi} \in K$, then the solution u solves the original Mumford-Shah problem optimally. The corresponding vector field $\bar{\varphi}$ is called a *calibration*. This is easily verified, because for any function $v : \Omega \to \mathbb{R}$ which agrees with u on the boundary of Ω we have due to the divergence theorem:

$$E(v) = \sup_{\varphi \in K} \int_{\Omega \times \mathbb{R}} \varphi D\mathbf{1}_v$$

$$\geq \int_{\Omega \times \mathbb{R}} \bar{\varphi} D\mathbf{1}_v = \int_{\Omega \times \mathbb{R}} \bar{\varphi} D\mathbf{1}_u = E(u), \quad (6)$$

which implies the minimality of $E(u)$. For the sake of completeness let us quickly sketch the proof of Theorem 1. For simplicity, we will present the proof for the case $\lambda = 0$ only but an extension to $\lambda > 0$ is straight forward.

Proof. The proof start by expressing the integral on the right-hand side of (4) as a flux through the interface Γ_u representing the discontinuities of $\mathbf{1}_u$ – see Figure 2:

$$\int_{\Omega \times \mathbb{R}} \varphi D\mathbf{1}_u = \int_{\Gamma_u} \varphi \cdot \nu_{\Gamma_u} d\mathcal{H}^2, \quad (7)$$

where the normal ν_{Γ_u} on the interface Γ_u – see Figure 2 – is given by:

$$\nu_{\Gamma_u} = \begin{cases} \frac{(\nabla u, -1)}{\sqrt{|\nabla u|^2 + 1}}, & \text{if } u \in \Omega \backslash S_u \\ (\nu_u, 0), & \text{if } u \in S_u \end{cases}, \quad (8)$$

where ν_u denotes the unit normal on S_u pointing from the side of u^- to that of u^+. The flux through the interface Γ_u can therefore be written as:

$$\int_{\Gamma_u} \varphi \cdot \nu_{\Gamma_u} d\mathcal{H}^2 = \int_{\Omega \backslash S_u} (\varphi^x \cdot \nabla u - \varphi^t) dx + \int_{S_u} \left(\int_{u^-}^{u^+} \varphi^x dt \right) \nu_u d\mathcal{H}^1.$$

Thus the flux through Γ_u can be expressed as the sum of an integral over $\Omega \backslash S_u$ and integral over S_u. The two constraints on φ imposed in K assure that:

$$\varphi^x \cdot \nabla u - \varphi^t \leq |\nabla u|^2, \quad (9)$$

and that

$$\left| \int_{u_1}^{u_2} \varphi^x dt \right| \leq \nu, \quad (10)$$

which implies that the flux through Γ_u provides a lower bound on the Mumford-Shah energy:

$$E(u) \geq \sup_{\varphi \in K} \int_{\Gamma_u} \varphi \cdot \nu_{\Gamma_u} d\mathcal{H}^2. \quad (11)$$

In addition, one can show (through a rather technical proof) that it is possible to build vector fields in K such that (9) and (10) are almost equalities, up to an arbitrarily small error, showing that the equality in (4) holds. □

In order to compute a minimizer of the Mumford-Shah functional, we substitute $\mathbf{1}_u$ in (4) by a generic function $v(x,t) : \Omega \times \mathbb{R} \to [0,1]$ which satisfies

$$\lim_{t \to -\infty} v(x,t) = 1 , \quad \lim_{t \to +\infty} v(x,t) = 0 . \quad (12)$$

Hence, we are going to face the following optimization problem:

$$\min_v \left\{ \mathcal{E}(v) = \sup_{\varphi \in K} \int_{\Omega \times \mathbb{R}} \varphi Dv \right\}, \quad (13)$$

which clearly poses a convex optimization problem. However, the crucial question remains, whether an optimal pair (v^*, φ^*) of (13) also admits a global minimizer of the Mumford-Shah functional? This would be true if one could proof a co-area formula [11]

$$\mathcal{E}(v) = \int_0^1 \mathcal{E}(\mathbf{1}_{\{v > s\}}) ds. \quad (14)$$

Unfortunately, (14) does not hold in the case of the Mumford-Shah functional. On the other hand, this formula

already suggests that if v^* is binary, (14) is fulfilled and we have found a global minimizer of the Mumford-Shah functional. Let us finally mention that for functionals which are convex in Du, the co-area formula holds. For example, in our setting, this is the case for $\nu = +\infty$.

3. Numerical Algorithms

In this section we present numerical algorithms to compute the solution of the relaxed problem (13).

3.1. Discrete Setting

For notational simplicity we may assume that $\Omega = [0,1]^2$ and $u(x) \in [0,1]$. That is, the subgraph of the function u is defined in the unit cube $[0,1]^3$. For discretization, we use a regular $(N \times N) \times M$ pixel grid defined by

$$\mathcal{G} = \Big\{ (i\Delta_x, j\Delta_x, k\Delta_t) :$$
$$i, j = 1, 2, \ldots, N, \ k = 1, 2, \ldots, M \Big\}, \quad (15)$$

where (i, j, k), are the indices of the discrete locations on the grid and $\Delta_x = 1/N$ and $\Delta_t = 1/M$ denote the discretization widths of the domain and the codomain of u.

Next, let us introduce the variables $x \in X : \mathcal{G} \to \mathbb{R}$ and $y \in Y : \mathcal{G} \to \mathbb{R}^3$ which are the discrete versions of v and ϕ in (13). In the following we will treat x as the primal variable and y as the corresponding dual variable. Then we define a discrete version of (13):

$$\min_{x \in C} \max_{y \in K} \langle Ax, y \rangle \ , \quad (16)$$

where the linear operator A resembles the discrete gradient operator. For discretization, we use simple forward differences with Neumann boundary conditions. Furthermore, the convex set $C \subseteq X$ is given by

$$C = \Big\{ x \in X : x(i,j,k) \in [0,1] \Big\} \ , \quad (17)$$

and additionally satisfies the boundary constraints $x(i,j,1) = 1$ and $x(i,j,M) = 0$ to account for the limits given in (12). Finally, the convex set $K \subset Y$ is given by

$$K = \Big\{ y = (y^1, y^2, y^3)^T \in Y :$$
$$y^3(i,j,k) \geq \frac{y^1(i,j,k)^2 + y^2(i,j,k)^2}{4} - \lambda(\frac{k}{L} - f(i,j))^2,$$
$$\left| \sum_{k_1 \leq k \leq k_2} (y^1(i,j,k), y^2(i,j,k))^T \right| \leq \nu \Big\} \ . \quad (18)$$

In the next section we will present a novel globally convergent first order primal-dual algorithm to compute the solution of (16). It will turn out that the proposed algorithm is efficient in terms of computational complexity and outperforms existing methods.

3.2. A Fast Primal-Dual Algorithm

The optimization problem (16) poses a classical saddle-point problem. The algorithm we propose here is a variant of an old algorithm proposed by Popov in [18]. We have modified it in order to reduce the number of projections, which is the most costly step in our case. Our scheme is also related to a more simple primal-dual projected gradient ascend/descend scheme which has been used in [21], but for which convergence has not been proven so far.

The algorithm we are going to describe here applies for a larger class of problems including pointwise linear terms $\langle g, x \rangle$ and $\langle h, y \rangle$. Problem (16) can obtained by simply setting $g = h = 0$. However, let us turn back to the more general problem, which is of the form

$$\min_{x \in C} \max_{y \in K} \langle Ax, y \rangle + \langle g, x \rangle - \langle h, y \rangle \ , \quad (19)$$

where $C \subseteq X$, $K \subseteq Y$ are closed, convex sets, $A : X \to Y$ is a continuous linear operator with norm L. X and Y are finite-dimensional spaces, although the proof also holds for Hilbert spaces: however, in this case, the problem is purely academic since in most interesting cases A would not be bounded. We assume the problem has at least a solution $(\hat{x}, \hat{y}) \in C \times K$, and in particular, we have for any $(x, y) \in C \times K$,

$$\langle A^*y + g, x - \hat{x} \rangle - \langle Ax - h, y - \hat{y} \rangle \geq 0 \,, \quad (20)$$

where A^* denotes the adjoint operator of A.

Our algorithm is as follow: we choose $(x^0, y^0) \in C \times K$ and let $\bar{x}^0 = x^0$. We choose two time-steps $\tau, \sigma > 0$. Then, we let for each $n \geq 0$

$$\begin{cases} y^{n+1} = \Pi_K(y^n + \sigma(A\bar{x}^n - h)) \\ x^{n+1} = \Pi_C(x^n - \tau(A^*y^{n+1} + g)) \\ \bar{x}^{n+1} = 2x^{n+1} - x^n \end{cases} \quad (21)$$

(observe that $\bar{x}^{n+1}$ might not be in C, if it is not a linear space). Then, we have the following theorem:

Theorem 2. *Choose τ, σ such that $\tau\sigma L^2 < 1$. Then, as $n \to \infty$, $(x^n, y^n) \to (x^*, y^*)$ which solves (19).*

The proof of Theorem 2 is presented in Appendix A.

Let us now apply the proposed algorithm to problem (16). The Lipschitz parameter for the linear operator A used in 16) is computed as $L = \sqrt{12}$ (assuming that $\Delta_x = \Delta_t = 1$). According to Theorem 2 we choose $\tau = \sigma = \frac{1}{\sqrt{12}}$.

In each iteration of our primal-dual algorithm we have to perform Euclidean projections of x and y onto the feasible sets C and K. The projection of x onto C can be achieved
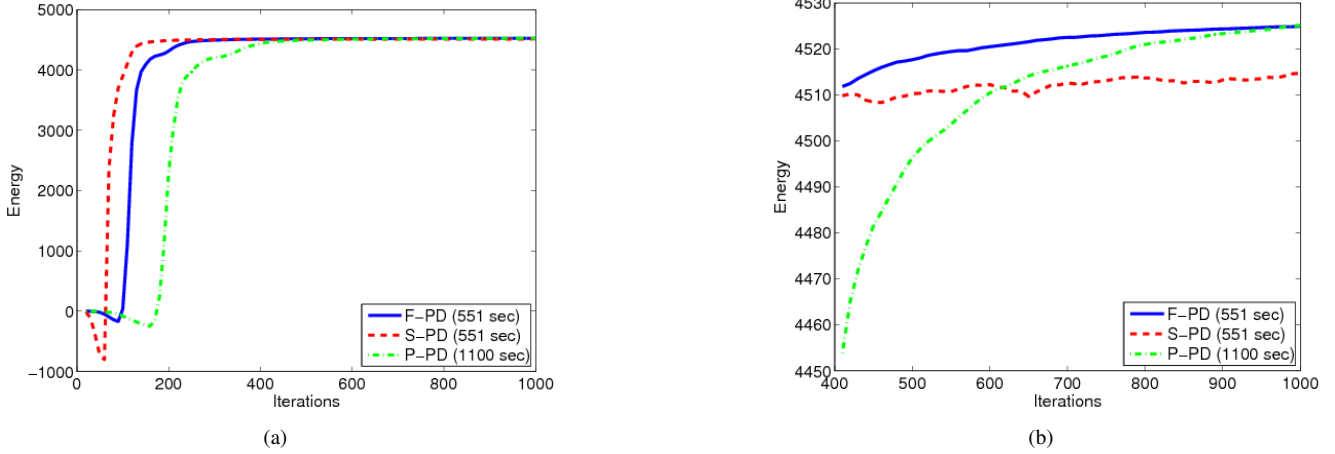
(a)

(b)

Figure 3. Comparison of the convergence behavior of F-PD, S-PD and P-PD using the synthetic test image of Figure 4. (a) shows the global convergence behavior along 1000 iterations and (b) shows the last 600 iterations. Note that S-PD is very fast in the beginning but shows a bad asymptotic convergence behavior. F-PD and P-PD show an iproved asymptotic convergence behavior, whereas the computational complexity of the proposed algorithm (F-PD) is about the half of P-PD.
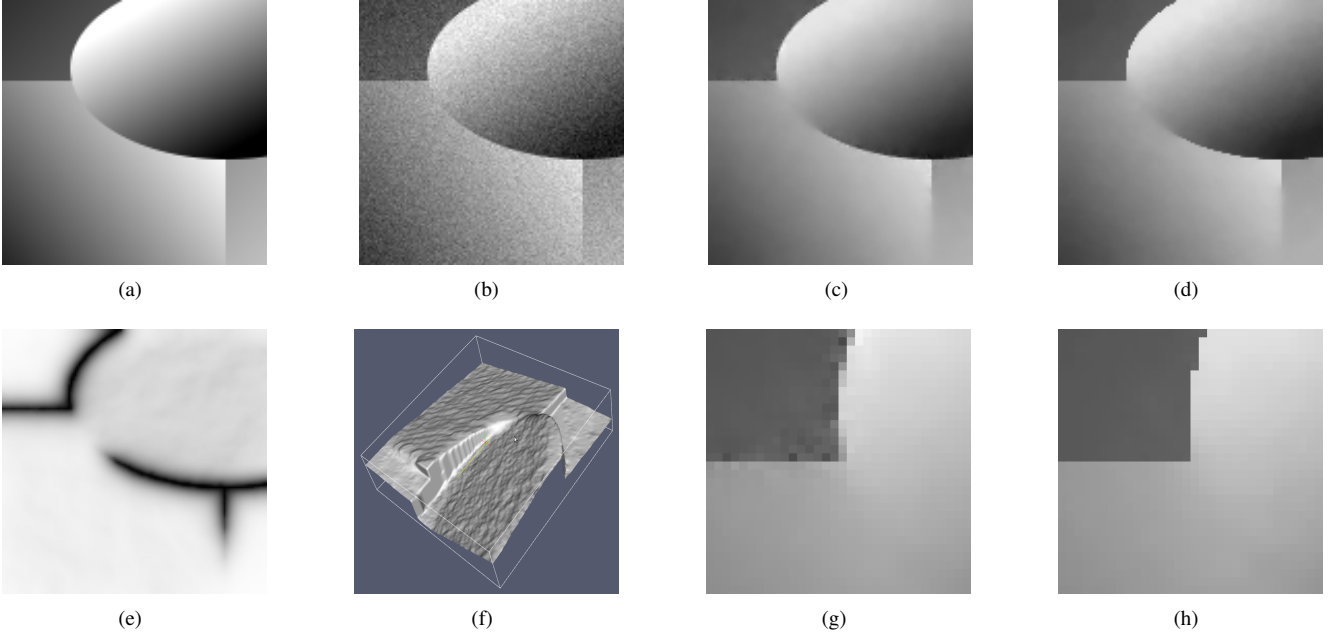

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 4. Piecewise smooth approximation of a synthetic test image of size 128×128 shown in (a). (b) shows the image degraded by 5% Gaussian noise. (c) shows the result of the phase field approach of Ambrosio and Tortorelli, the corresponding phase field is depicted in (e). (d) shows the result using the proposed approach, (f) shows the 0.5-isosurface rendering of the corresponding higher-dimensional solution. (g) and (h) show zoomings of (c) and (d). Note the artifacts of the method of Ambrosio and Tortorelli near discontinuities.

by a simple truncation operation $x = \min\{1, \max\{0, x\}\}$. The projection of y onto K is more involved since it takes into account non-local constraints. We therefore use Dykstra's iterative projection algorithm [6] to compute the projection. In the inner loop of Dykstra's algorithm we use Newton's algorithm to perform the projection onto the parabola as imposed by the first constraint in (18) an a simple soft shrinkage schemes to account for the non-local con-

straints in (18).

Figure 3 shows a comparison of our proposed fast primal-dual algorithm (F-PD) to the simple primal-dual algorithm (S-PD) of [21] and Popov's primal-dual algorithm (P-PD) [18]. The comparison is taken using the synthetic test image of Figure 4.

One can observe that S-PD is very fast during the first iterations, but after approximately 400 iterations, F-PD out-
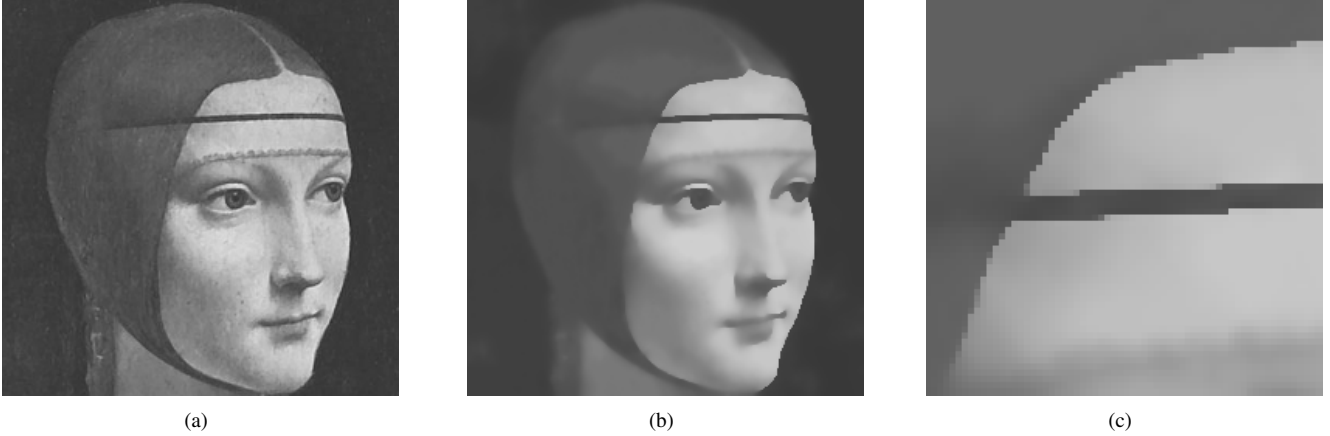
(a)	(b)	(c)

Figure 5. Piecewise smooth approximation of the drawing "La dama dell'ermellino" of Leonardo Da Vinci. (a) shows the 256×256 input image, (b) shows the piecewise smooth approximation using the proposed approach and (c) depicts a zooming which shows that the proposed method is able to preserve fine scale structures with sharp discontinuities.

performs S-PD and also shows an improved asymptotic behavior (see the zooming). It is important to note that F-PD has almost the same computational complexity as S-PD since it needs the same number of projections, which is the most costly part. P-PD is very slow in the beginning but after approximately 600 iterations, it also outperforms S-PD. After 1000 iterations P-PD reaches the same dual energy but we point out that P-PD needs twice as much projections as our algorithm and is therefore approximately two times slower.

3.3. Parallel Implementation

Our proposed primal-dual algorithm can be effectively parallelized on graphics processing units (GPUs). We implemented our algorithm on the GPU using the Nvidia CUDA framework. The algorithm are executed on a Tesla C1060 GPU running a 64 Bit Linux system. The speedup compared to a C/C++ implementation is about a factor of 30. Typical execution times for 128×128 images with the solution discretized on 32 levels are in the order of 600 seconds (see also Figure 3).

4. Experimental Results

In this section we provide experimental results of our algorithm using synthetic and real images. In all experiments, the proposed F-PD algorithm was iterated until the change of energy was below a certain threshold. The final solution $u(x)$ was computed out of the higher-dimensional function $v(x, t)$ be extracting the 0.5-isosurface.

4.1. Synthetic example

In our first experiment, we apply our algorithm for computing a piecewise smooth approximation of a synthetic test image. The clean image has been degraded by adding 5% Gaussian noise. Our algorithm was executed by discretizing the solution u on $M = 32$ levels. Somehow surprisingly, this rather coarse discretization of u already leads to high quality solutions. The parameters where set to $\nu = 5$ and $\lambda = 0.1$. Figure 4 shows a comparison of the proposed method to the phase field approximation of Ambrosio and Tortorelli (AT). One can observe that our method leads to smooth regions and sharp boundaries whereas the method of Ambrosio and Tortorelli shows artifacts near discontinuities (see also the zoomings).

4.2. Natural image

Figure 5 shows a piecewise smooth approximation of the drawing "La dama dell'ermellino" of Leonardo Da Vinci. The solution u was discretized using $M = 32$ and the parameters were set to $\nu = 5$ and $\lambda = 0.1$. This example shows that the proposed method is able to preserve fine-scale structures with sharp discontinuities.

4.3. Crack Tip

One central problem of the Mumford-Shah functional is the proof of optimality and existence of the so-called crack-tip problem, which plays a central role in the modeling of open boundaries in segmentation problems and is also important for the study of cracks in fracture mechanics. In order to compute the crack-tip example we generated a synthetic input image with the function $I(x, y) = \sqrt{r(x, y)} \sin(\theta(x, y)/2)$, where $r(x, y)$ is the Euclidean distance of a point (x, y) to the image center and $\theta(x, y)$ is the angle of the point (x, y) to the horizontal line. The function u was discretized using $M = 64$ levels and the parameter ν was set to 12. The parameter λ was set to 0 inside a disk-like region and set to $+\infty$ outside. Hence the solution
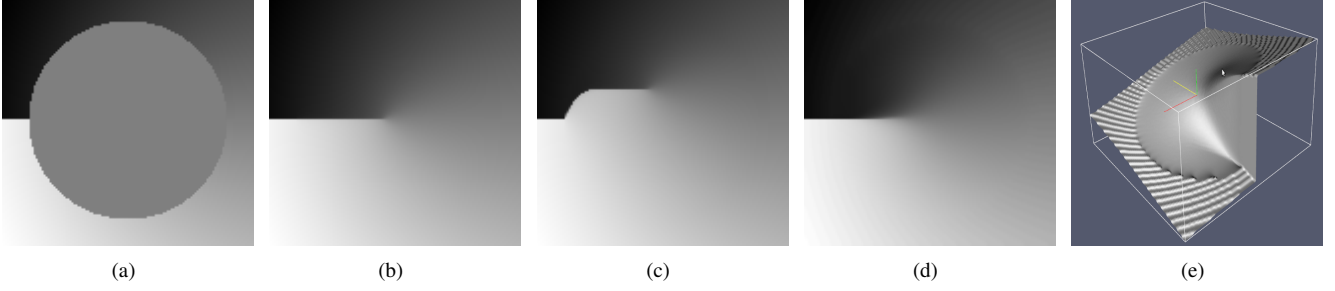
Figure 6. Computing the solution of the crack tip problem. (a) shows the 127×127 input image. The gray disk depicts the region where the parameter λ was set to 0. (b) shows the result of the approach of Ambrosio and Tortorelli from an initial guess close to the true solution and (c) shows the solution obtained from a bad initial guess. (d) shows the result of the proposed algorithm and (e) shows the 0.5-isosurface rendering of the corresponding higher-dimensional solution. Note that our method gives a good approximation to the crack-tip problem, which is independent from any initialization whereas the method of Ambrosio and Tortorelli heavily relies on the initial guess.

$u(x, y)$ takes exactly the same values as $I(x, y)$ outside the disk and minimizes the Mumford-Shah energy inside the disk. Figure 6 shows a comparison of our convex method with the non convex phase field approach of Ambrosio and Tortorelli. One can see that the method of Ambrosio and Tortorelli leads to good results only when providing a good initial guess. In contrast, since being a convex method, our method delivers a good approximation of the crack-tip independent of any initialization.

5. Conclusion

We proposed an algorithm for minimizing a convex relaxation of the piecewise smooth Mumford-Shah functional. The convex formulation is obtained by functional lifting in higher dimension and convex relaxation. The solution of the relaxed problem amounts to finding saddle points of a convex-concave functional. We propose to solve this by a novel primal-dual projected gradient descent/ascend algorithm for which we prove convergence. To the best of our knowledge, this is the first algorithm to minimize the piecewise smooth Mumford-Shah functional in a manner that is independent of initialization. In experimental comparisons on synthetic and real images we demonstrate that it it better reconstructs piecewise smooth signals than the commonly used approach of Ambrosio and Tortorelli.

Appendix A

Proof. We closely follow the proof in [18]. First, we observe that for any $u \in Y$ and $v \in K$,

$$\|v - \Pi_K(u)\|^2 \leq \|v - u\|^2 - \|u - \Pi_K(u)\|^2 \quad (22)$$

so that, choosing $v = \hat{y}$ and $u = y^n + \sigma(A\bar{x}^n - h)$, we find

$$\|y^{n+1} - \hat{y}\|^2 \leq \|y^n - \hat{y}\|^2 - \|y^n - y^{n+1}\|^2 + 2\sigma \langle A\bar{x}^n - h, y^{n+1} - \hat{y} \rangle. \quad (23)$$

In the same way,

$$\|x^{n+1} - \hat{x}\|^2 \leq \|x^n - \hat{x}\|^2 - \|x^n - x^{n+1}\|^2 - 2\tau \langle A^* y^{n+1} + g, x^{n+1} - \hat{x} \rangle. \quad (24)$$

We let for each n, $D_n = \sigma\|x^n - \hat{x}\|^2 + \tau\|y^n - \hat{y}\|^2$. Multiplying (23) by τ, (24) by σ, and adding both we get

$$D_{n+1} \leq D_n - \tau\|y^n - y^{n+1}\|^2 - \sigma\|x^n - x^{n+1}\|^2$$
$$-2\sigma\tau\big(\langle A^* y^{n+1} + g, x^{n+1} - \hat{x} \rangle - \langle A\bar{x}^n - h, y^{n+1} - \hat{y} \rangle\big) \quad (25)$$

It follows from (20) that

$$\langle A^* y^{n+1} + g, x^{n+1} - \hat{x} \rangle - \langle A\bar{x}^n - h, y^{n+1} - \hat{y} \rangle$$
$$\geq \langle A(x^{n+1} - \bar{x}^n), y^{n+1} - \hat{y} \rangle$$
$$= \langle A(x^{n+1} - 2x^n + x^{n-1}), y^{n+1} - \hat{y} \rangle$$
$$= \langle A(x^{n+1} - x^n), y^{n+1} - \hat{y} \rangle - \langle A(x^n - x^{n-1}), y^n - \hat{y} \rangle$$
$$+ \langle A(x^n - x^{n-1}), y^n - y^{n+1} \rangle$$
$$\geq \langle A(x^{n+1} - x^n), y^{n+1} - \hat{y} \rangle - \langle A(x^n - x^{n-1}), y^n - \hat{y} \rangle$$
$$- L\|x^n - x^{n-1}\|\|y^n - y^{n+1}\| \quad (26)$$

We choose $M > N \geq 1$ and sum (25) from N to M, using (26) and $2ab \leq \delta a^2 + b^2/\delta$ for any a, b and any $\delta > 0$. It follows

$$D_{M+1} \leq$$

$$D_N - \tau \sum_{n=N}^{M} \|y^n - y^{n+1}\|^2 - \sigma \sum_{n=N}^{M} \|x^n - x^{n+1}\|^2$$
$$-2\tau\sigma\Big(\langle A(x^{M+1} - x^M), y^{M+1} - \hat{y} \rangle$$
$$- \langle A(x^N - x^{N-1}), y^N - \hat{y} \rangle\Big)$$
$$+\tau\sigma L\Big(\delta \sum_{n=N-1}^{M-1} \|x^{n+1} - x^n\|^2 + \frac{1}{\delta} \sum_{n=N}^{M} \|y^n - y^{n+1}\|^2\Big).$$

Choosing $\delta = \sqrt{\sigma/\tau}$, so that $\sigma L/\delta = \tau L \delta = \sqrt{\sigma\tau}L < 1$, and $\delta' \in (\sigma L, 1/(\tau L))$, so that both $\tau' = \tau\sigma L/\delta' < \tau$ and $\tau\sigma L^2 < \tau L\delta' < 1$, it becomes

$$
\sigma\|x^{M+1} - \hat{x}\| + (\tau - \tau')\|y^{M+1} - \hat{y}\|^2
$$
$$
+ \tau(1 - \sqrt{\sigma\tau}L)\sum_{n=N}^{M}\|y^n - y^{n+1}\|^2
$$
$$
+ \sigma(1 - \sqrt{\sigma\tau}L)\sum_{n=N}^{M-1}\|x^n - x^{n+1}\|^2 \qquad (27)
$$
$$
+ \sigma(1 - \tau L\delta')\|x^M - x^{M+1}\|^2
$$
$$
\leq D_N + \tau^{1/2}\sigma^{3/2}L\|x^N - x^{N-1}\|^2
$$
$$
+ 2\tau\sigma\left\langle A(x^N - x^{N-1}), y^N - \hat{y}\right\rangle.
$$

We deduce that both $\|x^{M+1} - \hat{x}\|$ and $\|y^{M+1} - \hat{y}\|$ are bounded, and, as well, that both $\|y^{M+1} - y^M\|$ and $\|x^{M+1} - x^M\|$ go to zero as $M \to \infty$.

Hence, there is a subsequence (x^{M_k}, y^{M_k}) which converge to some couple $(x^*, y^*) \in C \times K$, and, moreover, also x^{M_k+1} and $\bar{x}^{M_k}$ converge to x^* while y^{M_k+1} goes to y^*. Passing in the limit in (21) we find that

$$
\begin{cases}
y^* = \Pi_K(y^* + \sigma(Ax^* - h)) \\
x^* = \Pi_C(x^* - \tau(A^*y^* + g)),
\end{cases}
$$

meaning that (x^*, y^*) is a solution of (19). Hence, we may choose $\hat{x} = x^*$, $\hat{y} = y^*$, and $N = M_k$ in (27): if k is large enough this will make the right-hand side of (27) arbitrarily small, hence for $M > M_k$ we get a bound on $\|x^{M+1} - \hat{x}\|$ and $\|y^{M+1} - \hat{y}\|$ which can be as small as we want. It follows that $(x^M, y^M) \to (x^*, y^*)$ as $M \to \infty$. $\qquad\square$

References

[1] G. Alberti, G. Bouchitté, and G. Dal Maso. The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calc. Var. Partial Differential Equations*, 16(3):299–333, 2003.

[2] L. Ambrosio. Variational problems in SBV and image segmentation. *Acta Appl. Math.*, 17:1–40, 1989.

[3] L. Ambrosio and V. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Comm. Pure Appl. Math.*, 43:999–1036, 1990.

[4] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[5] G. Bouchitté. Recent convexity arguments in the calculus of variations. (lecture notes from the 3rd Int. Summer School on the Calculus of Variations, Pisa), 1998.

[6] J. P. Boyle and R. Dykstra. A method for finding projections onto the intersection of convex sets in Hilbert spaces. In *Advances in order restricted statistical inference (Iowa City, Iowa, 1985)*, volume 37 of *Lecture Notes in Statist.*, pages 28–47. Springer, Berlin, 1986.

[7] X. Bresson, S. Esedoglu, P. Vandergheynst, J. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *J. Math. Im. Vis.*, 28(2):151–167, 2007.

[8] A. Chambolle. Finite differences discretization of the Mumford-Shah functional. *Math. Modelling and Numerical Analysis*, 33:261–288, 1999.

[9] T. Chan, S. Esedoḡlu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.

[10] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion Snakes: Introducing statistical shape knowledge into the Mumford–Shah functional. *Int. J. of Computer Vision*, 50(3):295–313, 2002.

[11] H. Federer. *Geometric Measure Theory*. Springer, 1969.

[12] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 6(6):721–741, 1984.

[13] M. Gobbino. Finite difference approximation of the Mumford-Shah functional. *Comm. Pure Appl. Math.*, 51(2):197–228, 1998.

[14] L. Grady and C. Alvino. Reformulating and optimizing the mumford-shah functional on a graph - a faster, lower energy solution. In *European Conference on Computer Vision (ECCV)*, pages 248–261, 2008.

[15] M. Morini and M. Negri. Mumford Shah functional as Γ-limit of discrete Perona-Malik energies. *Math. Models and Methods in Applied Sciences*, 13:785–805, 2003.

[16] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[17] T. Pock, D. Cremers, A. Chambolle, and H. Bischof. A convex relaxation approach for minimizing the Potts model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. to appear.

[18] L. Popov. A modification of the arrow-hurwicz method for search of saddle points. *Mathematical Notes*, 28(5):845–848, 1980.

[19] A. Tsai, A. J. Yezzi, and A. S. Willsky. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Trans. on Image Processing*, 10(8):1169–1186, 2001.

[20] L. Vese and T. Chan. A multiphase level set framework for image processing using the Mumford–Shah functional. *Int. J. of Computer Vision*, 50(3):271–293, 2002.

[21] M. Zhu and T. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. Technical Report 08-34, UCLA CAM Reports, 2008.