

Variational Segmentation for Color images

Alexander Brook (sova@cs.technion.ac.il)

Dept. of Mathematics, Technion, Israel

Ron Kimmel

Dept. of Computer Science, Technion, Israel

Nir A. Sochen

Dept. of Applied Mathematics, Tel-Aviv University, Israel

Abstract. We propose and analyze extensions of the Mumford-Shah functional for the color images. Our main motivation is the concept of images as surfaces. We also review most of the relevant theoretical background and computer vision literature.

Keywords: color, Mumford-Shah functional, segmentation, variational methods

1. Introduction

1.1. IMAGE SEGMENTATION—THE PROBLEM

For a long time now the vision problems have been subdivided into three classes: low-, intermediate- and high-level, each concerned with its own level of image description. On the low level we seek description in terms of edges, ridges, specularities, fragments of lines or circles; on the intermediate level—in terms of objects, their geometry, background, occlusions, etc; on the high level we expect to recognize the objects and give a full three-dimensional scene description. Typical problems on each level are edge detection, segmentation, and object recognition, respectively. This subdivision probably stems from David Marr’s model of vision [53].

The vision process, especially at the lower levels, is usually considered to be bottom-up. The importance of top-down feedbacks both within the same level and between levels is recognized sometimes, but is usually neglected in practice. Nevertheless, a thoughtful look at a few pictures readily convinces that in many cases edge detection, segmentation, three-dimensional reconstruction and image understanding are impossible without each other. Neuroanatomical [43] and psychophysical [48] evidence suggests that in primates these processes influence each other, and take place simultaneously or at least overlap in time.

A “neighbor” field of image processing poses other problems: image enhancement, restoration, compression—that may seem different and even unrelated. However in reality, when we have to deal with noisy,

blurred, distorted images, some restoration and enhancement are necessary before we can extract information from the image even at the low level. And vice versa: information obtained from vision algorithms (edges, segmentation) can help enormously to achieve good restoration and compression.

All this suggests that a segmentation or edge detection algorithm that incorporates as much of vision levels as possible and attempts restoration concurrently with segmentation should be worthwhile. Sadly, object recognition or full 3-dimensional description from a single image are very difficult without some *a priori* knowledge. So, we are left with edge detection, segmentation, and restoration. Variational formulation provides a framework that can integrate these problems and suggest algorithms for their solution.

1.2. IS IT WORTH IT?

The vast amount of existing algorithms for edge detection and segmentation compels to provide some justification, in addition to the general ideas above, before embarking upon developing yet another algorithm. Here we demonstrate how the most basic version of the Mumford-Shah functional (that will be the main theme of this work) can be used to improve drastically the performance of the Sobel edge detector.

Let us consider a circle (a color image in the range $[0, 1]^3$) and add to it Gaussian noise with zero mean and standard deviation 0.8 (Fig. 1b). The gradient magnitude (c) does not show clear maximum at the edge, and of course the edge detector performance is very poor (d). On the other hand, if we take the edge function produced by the implementation of the Mumford-Shah functional described in Section 4.1 (e), we see a clear edge, and the same edge detector performs very well (f).

1.3. THIS WORK

This work is an attempt to provide a general variational framework for color (or general vectorial) images, generalizing the Mumford-Shah functional. We also give a review of the variational methods of segmentation and edge detection. The initial intent was to provide a theoretical background for the model proposed and implemented in [45]. These plans, however, changed due to the need to find the right balance of model plausibility, quality of numerical results and theoretical validity.

One important issue is absent from this work, in spite of it being closely related to the theme and often leading to similar formulations and results. We do not consider at all the Markov Random Fields point of view on the segmentation problem. It should be noted that a great

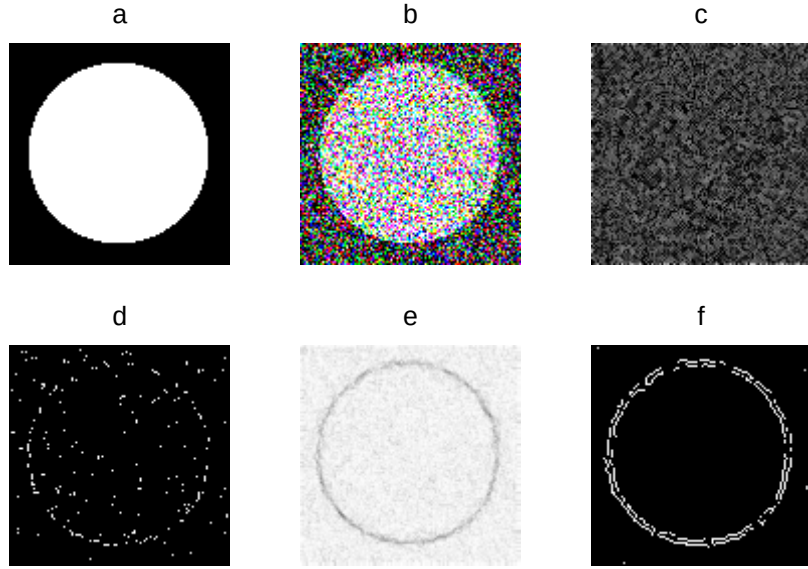


Figure 1. The noisy circle experiment.

deal of research was done in this field, and there are some promising results (see, e.g., [51]).

Another omission that we have to mention is the functionals depending on the second-order derivatives, as the weak plate model of [12]. These functionals were recently studied in [3], where elliptic approximations are provided and implemented numerically.

This work is organized as follows: in Section 2 we review the relevant work on variational segmentation and color edge detection. Section 3 offers a summary of the theory of the Mumford-Shah functional and of numerical minimization methods devised for this functional. We propose some generalizations of the Mumford-Shah functional in Section 4, discuss them and show some numerical results. Our conclusions and some directions for further research are presented in Section 5.

A few words on the notation. When a sentence defines a new term, this term is shown *slanted*. The norm $|\cdot|$ is the usual Euclidean norm of *any* object: a number, a vector, or a matrix. In particular, for a function $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$ we put $|\nabla u| = (\sum (\frac{\partial u^i}{\partial x_j})^2)^{1/2}$ (also called Hilbert-Schmidt or Frobenius norm of the ∇u matrix). $\mathcal{L}^n$ is the Lebesgue measure on $\mathbb{R}^n$.

$\mathcal{H}^{n-1}$ is the $(n-1)$ -dimensional *Hausdorff measure*, which is a generalization of the area of a submanifold. For any $A \subset \mathbb{R}^n$ we take a

covering $\{S_j\}$ of A (that is, $A \subset \cup S_j$); the *diameter of the covering* is defined to be $\lambda(\{S_j\}) = \sup_j \text{diam } S_j$. Then

$$\mathcal{H}^m(A) = \lim_{\delta \rightarrow 0} \inf_{\lambda(\{S_j\}) < \delta} \sum_j \omega_m \left(\frac{\text{diam } S_j}{2} \right)^m,$$

where ω_m is the (Lebesgue) volume of a unit ball in $\mathbb{R}^m$.

2. Image segmentation—a biased review

2.1. VARIATIONAL SEGMENTATION

We consider images as functions from a domain in $\mathbb{R}^2$ into some set, that will be called the *feature space*. When needed, we suppose that the domain is $[0, 1]^2$. Some examples of feature spaces are

- an interval, e.g. $[0, 255]$ or $[0, \infty)$, for gray-level images;
- a subset of $\mathbb{R}^3$, e.g. $[0, 1]^3$ or $S^1 \times [0, 1]^2$, for color images in RGB or HSV;
- $\mathbb{R}^2$ or $S^1 \times \mathbb{R}_+$ for a movement field.

2.1.1. Regularization

Regularization is among the well-established techniques of image restoration. A standard model of image acquisition is given by

$$g = Lf + e. \tag{1}$$

Here

- L is a linear operator representing the influence of the optical system, usually a Hilbert-Schmidt operator $Lf(x) = \int K(x, y)f(y)dy$.¹ The effect of L is a blur; when it is space-independent we have $K(x, y) = \bar{K}(x - y)$, meaning L is a convolution.
- e is an additive noise introduced by the recording system, which is usually assumed to be random, sometimes with known mean value and variance σ^2 .

In order to restore the image we have to solve Eq. (1) for f , which is usually an ill-posed problem; the main reason for the ill-posedness

¹ This assumption makes (1) a Fredholm integral equation of the first kind.

is that L is compact, and so L^{-1} is unbounded even when it exists. Another problem is that e is unknown.

We expect from (1) that $\|Lf - g\|^2 \approx \sigma^2$. To compensate for the loss of information in (1) we must use our *a priori* knowledge about the image, usually in the form of a smoothness assumption. That is, we assume that $\int |\nabla f|^2$, or $\int \|Df\|^2$ (for some differential operator D) is small. So, we arrive at the problem

$$\min_f \int \|Df\|^2 \quad \text{subject to} \quad \|Lf - g\|^2 = \sigma^2.$$

Using Lagrange multipliers, we arrive at an equivalent problem of

$$\min_f \left(\|Lf - g\|^2 + \alpha \int \|Df\|^2 \right).$$

The mathematical framework for this kind of argument was provided by Tikhonov in 1963 (see [8]). Consider an operator $L : F \rightarrow G$ on Hilbert spaces, and a closed convex subset $E \subset F$ representing *a priori* constraints. Given L_h and g_δ such that $\|L - L_h\| \leq h$ and $\|g - g_\delta\| \leq \delta$, the problem is to construct an approximate solution of $Lf = g$, $f \in E$. Tikhonov proposed to do it by minimizing the functional $\|L_h f - g_\delta\|_G^2 + \alpha \|f\|_F^2$, where $\alpha > 0$ is a parameter (to be chosen separately). This problem is well-posed. Applying this to our problem, we might set $F = W^{1,2}$, $G = L^2$, $\delta = \sigma$ and α can be chosen to ensure $\|Lf - g_\delta\|_G^2 = \delta^2$.

Frequently in applications $L_h = I$, thus the functional becomes $\|f - g_\delta\|_G^2 + \alpha \|f\|_F^2$, in our case $\int ((f - g)^2 + \alpha |\nabla f|^2)$.

2.1.2. Total variation methods

The presence of the term $\|f\|_F^2 = |\nabla f|^2$ in Tikhonov regularization often leads to smoothing and blurring of the edges; the Dirichlet functional $\int |\nabla f|^2$ “prefers” smooth gradients and “punishes” steep edges (e.g. the characteristic function of a unit ball in $\mathbb{R}^n$ is not in $W^{1,2}(\mathbb{R}^n)$). However, most images contain steep edges, which provide very important perceptual clues, and we would like to recover these steep edges during the reconstruction. The problem is that these edges are represented by discontinuities in the corresponding functions.

A solution was proposed in [60, 63], based on shock capturing numerical methods from fluid mechanics, and suggesting to minimize $\int ((f - g)^2 + \alpha |\nabla f|)$; the term $\int |\nabla f|$ is called *total variation* of f .² This functional allows discontinuities in the object function, and is

² This is only one of the many possible generalizations of total variation to the functions of several variables. It is sometimes referred to as total variation in the Tonelli’s sense.

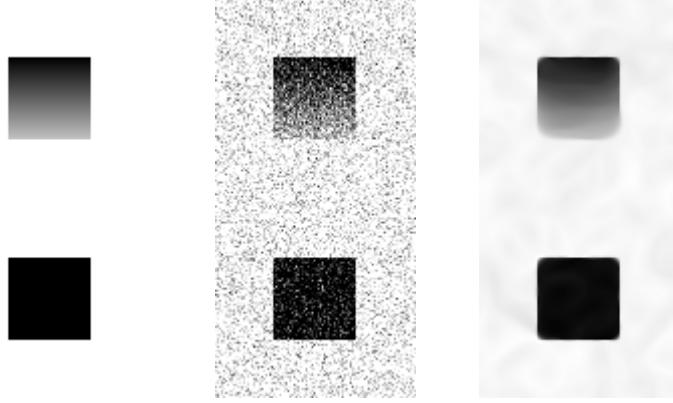


Figure 2. An example of total variation restoration. Note the good restoration of sharp edges and appearance of staircasing.

usually minimized over the space of functions of bounded variation (to be defined later). Total variation methods were extensively studied during the last ten years, both theoretically and practically; in particular, well-posedness is shown in [22]. An extensive bibliography can be found in [78, ch. 1.7].

The main drawbacks of the total variation reconstruction are

1. The integrand $|\nabla f|$ is not differentiable. The standard method to overcome this is to replace it with $\sqrt{|\nabla f|^2 + \beta}$, where $\beta > 0$ is a small parameter. Even then the resulting Euler-Lagrange equation is nonlinear and demands sophisticated numerical methods.
2. Although allowing discontinuities in the object function, total variation functional still “punishes” each discontinuity, proportionally to the height of the jump. An ideal image restoration functional should not punish large jumps (probable edges), definitely not more than small ones (probable noise). In this respect, total variation exhibits behavior which is opposite to our expectations.
3. Very strong “staircasing” effect on noisy images which are far from being piecewise constant. To overcome this, we can use total variation restoration near edges and $W^{1,1}$ restoration in the smooth regions; two different methods based on this idea were suggested and implemented in [22] and [14].

2.1.3. Mumford-Shah functional

In [57] Mumford and Shah suggested segmenting an image by minimizing a functional of the form $\int_{\Omega \setminus K} (|\nabla f|^2 + \alpha |f - g|^2) + \beta \text{length}(K)$,

where K is the union of edges in the image. This choice is suggested by modeling images as piecewise smooth functions. The image is supposed to consist of a number of regions with slow and smooth changes within each region. Across the boundaries between these regions the changes may be abrupt. Of course, we must also suppose that these boundaries are “nice” (a union of smooth curves of small length, for example).

The functional consists of three terms:

- the *smoothing term* $\int_{\Omega \setminus K} |\nabla f|^2$, which should be small, if f is changing slowly within regions;
- the *fidelity term* $\int_{\Omega \setminus K} |f - g|^2$, that controls how close the smoothed image should be to the input;
- the length of the edges $\text{length}(K)$, that must be kept small to prevent the edges from filling up the whole image.

A justification of this model from the statistical point of view is given in [56]. Morel and Solimini in their book [55] show that many other segmentation models are particular cases of the Mumford-Shah functional or are closely related to it.

The minimization of the Mumford-Shah functional poses a difficult problem, both theoretical and numerical, because it contains both area and length terms and is minimized with respect to two variables: a function $f : \Omega \rightarrow \mathbb{R}$ and a set $K \subset \Omega$. This kind of functionals was introduced in [33]. In [32] De Giorgi introduced the name “free discontinuity problems”, referring to his idea of representing K as the set of jump points of f .

2.2. COLOR IMAGE SEGMENTATION

There are numerous sources of vectorial images, i.e., those with feature space of dimension higher than one. The most obvious, widespread and important are color images. Other examples include multi-modal medical images, satellite images taken at a set of wavelengths, spectral imaging and others. Sometimes vectorial images are derived from scalar images, e.g. decompositions of texture data with respect to some basis or a collection of image derivatives of some orders. Optical flow is yet another source of vectorial images.

We will work with the usual color images, for simplicity in the RGB color space. The easiest way to extend image processing methods onto color images is the channel-by-channel processing. Unfortunately, it is frequently inadequate for segmentation purposes. Some edges may be strong only in one channel and remain undetected. Since any smoothing unavoidably shifts the edges slightly, edges that are strong in two or

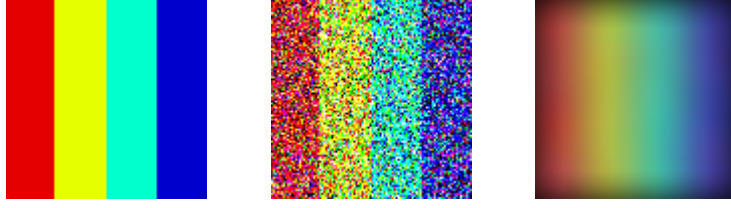


Figure 3. Channel-by-channel color image restoration. The restoration was performed by a convolution with a Gaussian. Note the appearance of a green stripe, though there is no green in the original image.

three channels will produce thin stripes of spurious colors (see Figure 3). Thus, some coupling between the channels is needed.

The usual mean of providing coupling is by defining a suitable *edge indicator function* e , that is supposed to be small in the smooth parts of the image and large in the vicinity of an edge. A typical example is $e(x) = |\nabla f(x)|^2$, and the integral $\int e$ usually constitutes the smoothing term.

2.2.1. Images as manifolds

One of the promising frameworks to derive and justify edge indicators is to consider images as embedded manifolds and to look at the induced metric for qualitative measurements of image smoothness. This idea first appeared in [36]. In this article images are considered as functions $f : \mathbb{R}^2 \rightarrow \mathbb{R}^m$, thus defining a two-dimensional manifold in $\mathbb{R}^m$ with induced metric $g_{ij} = \left\langle \frac{\partial f}{\partial x_i}, \frac{\partial f}{\partial x_j} \right\rangle$ (under the assumption that $\text{rank} \frac{\partial f_i}{\partial x_j} = 2$ everywhere). The case $m = 1$ of a usual grey-level image makes no sense in this setting, however.

In the work [29] an edge indicator based on g is proposed and implemented: edges are where the greater eigenvalue of g has a maximum in the direction of the corresponding eigenvector. Also, noticing the problem when $m = 1$, the author suggests representing an image as a surface embedded in $\mathbb{R}^{m+2}$ given by $(x_1, x_2, f_1, \dots, f_m)$ as a “more consistent geometric interpretation”.

This interpretation was formulated in the most general way and implemented in [72]. The n -dimensional m -valued image is considered as an n -dimensional manifold in $(\mathbb{R}^{n+m}, h)$ given by

$$(x_1, \dots, x_n, f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)),$$

and $\sqrt{\det g}$, where g is the metric on the manifold induced by the metric h from $\mathbb{R}^{n+m}$, is taken to be the edge indicator function. The integral $\int \sqrt{\det g}$ gives the n -dimensional volume of the manifold, and

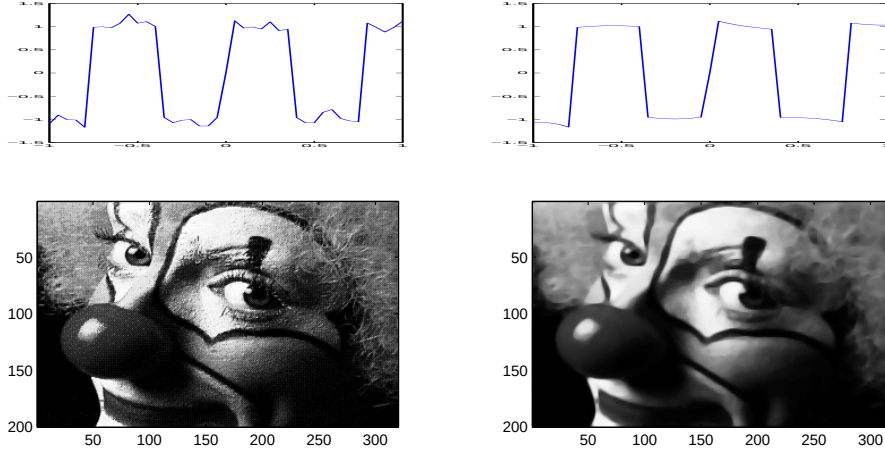


Figure 4. Examples of Beltrami flow restoration.

its minimization brings on a kind of non-isotropic diffusion, which the authors called the Beltrami flow. A similar flow was also independently suggested in [80].

As pointed out in [72, 80, 73], when implementing such a diffusion, one must decide what is the relationship between unit lengths along the x_i axes and along the f_j axes. It seems that no general principle exists to help in this decision. The significance of the ratio of the scales is discussed in detail in [72]. We will denote this coefficient by γ .

The simplest example of this framework is $f : \mathbb{R} \rightarrow \mathbb{R}$. The corresponding energy is just the length of the graph $(x, f(x))$, and its minimization is by the curvature flow.

In the case of gray-level images this framework was first introduced in [38]. Here the image is a surface in $\mathbb{R}^3$, the edge indicator is the area element $(1 + f_x^2 + f_y^2)^{1/2}$, and the flow is closely related to the mean curvature flow.

In a number of works (e.g. [23, 75, 46]) another problem is considered, leading to very similar equations. It is the problem of smoothing, scaling and segmenting an image in a “non-flat” feature space, like a circle, a sphere or a projective line.

2.2.2. Related work

Segmentation and restoration of vectorial images are not always a straightforward generalization of grey-level image segmentation, and different possibilities were offered and explored.

In [49] a texture image is decomposed using Gabor wavelets and represented as a *scalar* function $f(\sigma, \theta, x, y)$ depending (besides position)

on the frequency (σ) and direction (θ) of the wavelet. This function then serves as the argument for a Mumford-Shah type functional, but with quadruple integrals. It seems impossible to use this continuous formulation for cases where the feature space is finite-dimensional (like in color images) but the computer implementation (which uses 24-dimensional sampling, 3 frequencies and 8 orientations, of the feature space) can be used for all cases. An essential point in this paper is that the right choice of a norm on the feature space is important; the authors argue convincingly in favor of using L_2 and not the L_∞ .

The representation in [9] is again a Mumford-Shah type functional used to segment texture, but here the images are vectorial (texture decomposition using Zernike polynomials). The approximating function is supposed to be piecewise constant, which eliminates the smoothing term from the functional. Instead of the length of the discontinuity set its affine total variation is taken, thus making the segmentation affine-invariant. The implementation is by region growing.

Mumford-Shah functional is the model in [24], again with piecewise constant approximation. The coupling between channels is by the common edge set. Implementation is very different, though, using level sets to represent the edge.

The work [79] is concerned with diffusion rather than segmentation, but the framework is very similar. The Hilbert-Schmidt norm of the Jacobian squared ($|\nabla f|^2$) is proposed as diffusion coefficient, common for all channels. Also, a metric on the feature space is introduced: feature space of normalized image gradient is equipped with the S^1 metric.

Color images are regarded as functions $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ in [67, 66], carrying on the ideas from [29]: use the metric induced from the feature space to extract information on edges. In these works the expression $f(\lambda_+ - \lambda_-)$, where λ_+, λ_- are the eigenvalues of the metric, is used as a diffusion coefficient. CIELAB is used as a metric on the feature space (color).

Color TV (total variation) method, introduced in [13], proposes the use of the norm

$$\text{TV}_{n,m}(\Phi) = \left(\sum_{i=1}^m \left[\int_{\Omega} |\nabla \Phi^i| \right]^2 \right)^{1/2}$$

as a smoothing term for color image $\Phi : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, which is a generalization of total variation. This method performs very well for images with crisp edges, but has strong staircasing effect on smooth gradients. Besides, any TV method, given a “step” as an input will reduce its height; generally, we would like the opposite, so that the edges are enhanced.

A similar approach is taken in [71]: in a variational formulation for scalar image segmentation, the smoothing term is replaced by the $|\nabla f|$. The values are transformed to CIELAB.

A more general approach is adopted in [15]. The feature space is taken to be $[0, 1]^3$ with a Riemannian metric φ_{ij} . A distance $d(f(x), g(x))$ between the points of the two images is taken to be the geodesic distance, and the energy is

$$\int_{\Omega \setminus K} \left(d(f(x), g(x)) + \sum_{i,j} \varphi_{ij} \langle \nabla f^i, \nabla f^j \rangle \right) + \mathcal{H}^1(K).$$

3. Mumford-Shah functional

3.1. THEORY

3.1.1. *Weak formulation*

Initial formulation in [58] suggested minimizing

$$E(u, K) = \int_{\Omega \setminus K} (|\nabla u|^2 + \alpha |u - w|^2) + \beta \text{length}(K)$$

over $u \in C^1(\Omega \setminus K)$ and K a finite union of smooth arcs. Mumford and Shah conjectured that minimizers exists, and that there are three possible configurations for endpoints and crossings in K :

1. *triple points*, where three arcs meet at 120° ,
2. *boundary points*, where a curve meets the boundary perpendicularly,
3. *crack tips*, where a single arc ends without meeting others.

The core difficulty in proving this conjecture is that the functional is a sum of an area and a curvilinear integrals, and the curve of integration is one of the variables.

There is yet another problem, namely that the proposed domain of u and K is too restrictive and lacks some convenient properties (compactness, lower semicontinuity of the functionals in question). This one, however, is ordinary; instead of imposing that K is a finite union of smooth arcs, we should drop this requirement, and prove later that a minimizing K must be smooth. This also necessitates replacing $\text{length}(K)$ with something defined on non-smooth sets; the most natural replacement is $\mathcal{H}^1(K)$, the one-dimensional Hausdorff measure of K .

The crucial idea in overcoming the difficulties of interaction between the area and the length terms is to use a weak formulation of the problem. First, we let K be the set of jump points of u : $K = S_u$. The functional thus depends on u only. Second, we relax the functional in L^2 , that is, we consider

$$\bar{E}(u) = \inf \left\{ \liminf_{k \rightarrow \infty} E(u_k, S_{u_k}) : u_k \rightarrow u \text{ in } L^2, u_k \in C^1(\Omega \setminus \overline{S_{u_k}}), \mathcal{H}^1(\overline{S_{u_k}} \setminus S_{u_k}) = 0 \right\}.$$

It turns out (see [7]) that this functional has an integral representation

$$\bar{E}(u) = \int_{\Omega} (|\nabla u|^2 + \alpha|u - w|^2) + \beta \mathcal{H}^1(S_u)$$

and if $\bar{E}(u)$ is finite then $u \in \text{SBV}$, the space of special functions of bounded variation.

In this weak setting it was shown in [34] that $\bar{E}$ indeed has minimizers and that at least some of them are regular enough (with K closed and $u \in C^1(\Omega \setminus K)$). Actually, it was proven for the more general case of $\Omega \subset \mathbb{R}^n$, $n \geq 2$, and

$$F(u) = \int_{\Omega} (|\nabla u|^2 + \alpha|u - w|^2) + \beta \mathcal{H}^{n-1}(S_u).$$

3.1.2. BV functions

The question arises of the space on which to consider the functional F , in particular, how should S_u be defined. A class of functions is needed that is sufficiently regular for ∇u to exist a.e., but if we take too regular a class (like $W^{1,1}$) jump sets will be too small (of zero “length”). If we take a class too general, the jump set will be irregular, and we need it to be reasonably similar to a closed subset of finite length. It seems that the natural habitat for F is the space of functions of bounded variation. We will now define it and present some of its properties. The books [5, 39] can be consulted for a full treatment of the subject.

A function $u \in L^1(\Omega, \mathbb{R}^m)$ (for an open $\Omega \subset \mathbb{R}^n$) is of *bounded variation*, if

$$|Du|(\Omega) = \sup \left\{ \sum_{k=1}^m \int_{\Omega} u^k \operatorname{div} g^k \mid g^k \in C_0^1(\Omega, \mathbb{R}^n), \|g^k\|_{\infty} \leq 1 \right\} < +\infty.$$

Here $|Du|(\Omega)$ is the *(total) variation* of u in Ω . This definition agrees with one in Section 2.1.2 for smooth real functions.

We will say that u_0 is an *approximate limit* of u at x , and write $\operatorname{ap} \lim_{y \rightarrow x} u(y) = u_0$, if

$$\lim_{\rho \rightarrow 0} \rho^{-n} \int_{B_{\rho}(x)} |u(y) - u_0| dy = 0.$$

A function of bounded variation has an approximate limit $\mathcal{L}^n$ -a.e. The *approximate discontinuity set* S_u is the set of all points in Ω where u does not have an approximate limit. The set S_u is countably $\mathcal{H}^{n-1}$ -rectifiable, that is, it is (up to a $\mathcal{H}^{n-1}$ -negligible set) a union of countably many C^1 hypersurfaces. Hence, $\mathcal{H}^{n-1}$ -everywhere we can define a *unit normal* ν_u to S_u and *traces* of u on S_u by

$$u^\pm(x) = \lim_{\substack{y \rightarrow x \\ \pm \langle y-x, \nu_u(x) \rangle > 0}} u(y).$$

Functions of bounded variation are also $\mathcal{L}^n$ -a.e. *approximately differentiable*, that is, for $\mathcal{L}^n$ -a.e. $x \in \Omega$ there is a vector $\nabla u(x)$ such that

$$\lim_{y \rightarrow x} \frac{|u(y) - u(x) - \langle \nabla u(x), y - x \rangle|}{|y - x|} = 0.$$

So, we see that all the elements of F are well defined, and have their intended meaning.

3.1.3. SBV functions

However, it turns out that BV is too large for our purpose, since it contains function like the Cantor-Vitali function (Cantor's ladder), and every $w \in L^2$ can be approximated by such a function that has zero derivative and is continuous. Thus, $\inf_{\text{BV}} F = 0$, but this infimum is never reached.

The following result is classical for functions in $\text{BV}(\mathbb{R}, \mathbb{R})$: $f = f^a + f^j + f^c$, where f^a is absolutely continuous, f^j consists of a finite or a countable number of jumps, f^c is continuous and $\frac{d}{dx} f^c(x) = 0$ a.e. (in f^c 'c' if for Cantor). Sometimes this result is formulated in terms of measures: by the Radon-Nikodym theorem, $Du = D^a u + D^s u$, with $D^a u = \nabla u \mathcal{L}^1$ being the absolutely continuous part and D^s the singular part (with respect to $\mathcal{L}^1$). $D^s u$ can be further decomposed into the part supported on S_u (*the jump part*) and the rest (*the Cantor part*): $D^s u = D^j u + D^c u$, and thus $Du = D^a u + D^j u + D^c u$.

The latter decomposition generalizes to $\text{BV}(\mathbb{R}^n, \mathbb{R}^m)$. The functions without the Cantor part in their derivatives are the *special functions of bounded variation* and this space is denoted by $\text{SBV}(\mathbb{R}^n, \mathbb{R}^m)$. It is well suited for the study of functionals of the Mumford-Shah type.

We also note that $D^a u = \nabla u \mathcal{L}^n$ (here ∇u is the approximate gradient) and $D^j u = (u^+ - u^-) \otimes \nu_u \mathcal{H}^{n-1}|_{S_u}$.

For technical reasons, we will also need the following spaces: *generalized* BV (GBV) and *generalized* SBV (GSBV). A function u is in GBV(GSBV) if its truncations $\min(T, \max(T, u))$ are in $\text{BV}(\text{SBV})$ for

any $T > 0$. For the practical applications to images, where all functions are bounded, these distinctions are irrelevant.

3.1.4. *Known theoretical results*

Presently, the Mumford-Shah conjecture in its original form stands unproven, yet there are interesting and meaningful advancements (see the review [54] and the book [5]). For example, in [15] the conjecture is proved under the assumption that the number of connected components of K is bounded. A more general but less complete result appears in [5]: there exists a closed set $\Sigma \subset K$, $\mathcal{H}^{n-1}(\Sigma) = 0$ such that $K \setminus \Sigma$ is locally $C^{1,\nu}$ hypersurface for any $\nu < 1$ (in the case $n = 2$ also for $\nu = 1$).

Another result proved in [4], that is important for numerical approximations, is that a minimizer (u, K) of the Mumford-Shah functional can be approximated by pairs $(u_\varepsilon, K_\varepsilon)$ where K_ε is piecewise-smooth and $u_\varepsilon \in C^\infty(\Omega \setminus K)$, and such that $E(u, K) < E(u_\varepsilon, K_\varepsilon) + \varepsilon$. A similar result was proved in [37].

Besides the question of the regularity of minimizers, it has to be proven that the conjectured configuration are, indeed, minimizers (and the only ones). Considerable progress was made and is also reviewed in [5]. In a recent preprint [16] crack tip is shown to be a global minimizer.

An interesting and important limiting case of the Mumford-Shah functional is the problem

$$\bar{F} = \int_{\Omega} \alpha |u - w|^2 + \beta \mathcal{H}^{n-1}(S_u), \quad \nabla u = 0 \text{ on } \Omega \setminus S_u \quad (2)$$

of approximating g by a piecewise-constant function. For this functional, the Mumford-Shah conjecture was proved already in the original paper [58]; an elementary constructive proof can be found in [55]. Existence of minimizers for any $n \geq 2$ was shown in [28].

3.2. NUMERICAL APPROXIMATION

The main difficulty that hampers attempts to minimize the Mumford-Shah functional $E(u, K)$ numerically is the necessity to somehow store the set K , keep track of possible changes of its topology, and calculate its length. Also, the number of possible discontinuity sets is enormous even on a small grid.

We can, however, try to find another functional approximating the Mumford-Shah functional that will also be more amenable to numerical minimization. The framework for this kind of approximation is Γ -convergence, introduced in [35] (also see the book [31]).

3.2.1. Γ -convergence

A precursor of Γ -convergence was G-convergence, which was an attempt to invent a convergence for a certain type of differential equations that would imply convergence of their solutions. An example by De Giorgi showed that pointwise convergence was inappropriate for this purpose. Treating the differential equations as Euler-Lagrange equations of appropriate functionals, we can reformulate the problem as “find a convergence of functionals that implies convergence of minimizers”. Γ -convergence has this very nice property.

Consider a metric space (X, d) . A sequence of functionals $F_i : X \rightarrow \mathbb{R}_+$ is said to Γ -converge to $F : X \rightarrow \mathbb{R}_+$ ($\Gamma\text{-}\lim F_i = F$) if for any $f \in X$

1. $\forall f_i \rightarrow f : \liminf F_i(f_i) \geq F(f)$ (lower inequality);
2. $\exists f_i \rightarrow f : \limsup F_i(f_i) \leq F(f)$ (upper inequality, or existence of recovery sequence).

We can extend this definition to families of functionals depending on a continuous (real) parameter $\varepsilon \downarrow 0$, requiring convergence of F_{ε_i} to $F(x)$ on every sequence $\varepsilon_i \downarrow 0$.

It is important to notice that Γ -limit depends on what kind of convergence we have on X . Sometimes, to avoid ambiguities, it is designated as $\Gamma(X)$ - or $\Gamma(d)$ -limit.

We can take $X = \mathbb{R}$ to construct simple examples of Γ -convergence [31, 42]. For example (see Figure 5), $\Gamma\text{-}\lim_{n \rightarrow \infty} \sin nx = -1$, and $\Gamma\text{-}\lim_{n \rightarrow \infty} f_n = f$, where

$$f_n(x) = \begin{cases} -nx & 0 \leq x \leq 1/n, \\ nx - 2 & 1/n \leq x \leq 2/n, \\ 0 & \text{otherwise,} \end{cases} \quad f(x) = \begin{cases} -1 & x = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that in the first case there is no pointwise convergence and in the second case the pointwise limit of f_n is 0. In the first case every point is a limit point of a sequence of minimizers of $\sin nx$, and indeed, the limiting function has a minimum at every point. In the second case, $\arg \min f_n = 1/n \rightarrow 0$, and Γ -limit has a minimum at 0 with the correct value of -1 .

Some important properties of Γ -convergence:

- Γ -limit (if exists) is unique and lower semicontinuous. In particular, if $F_i \equiv F$ is a constant sequence, then $\Gamma\text{-}\lim F_i$ equals the lower semicontinuous envelope of F .
- If $\Gamma\text{-}\lim F_i = F$ and $G : X \rightarrow \mathbb{R}_+$ is continuous, then $\Gamma\text{-}\lim(F_i + G) = F + G$.

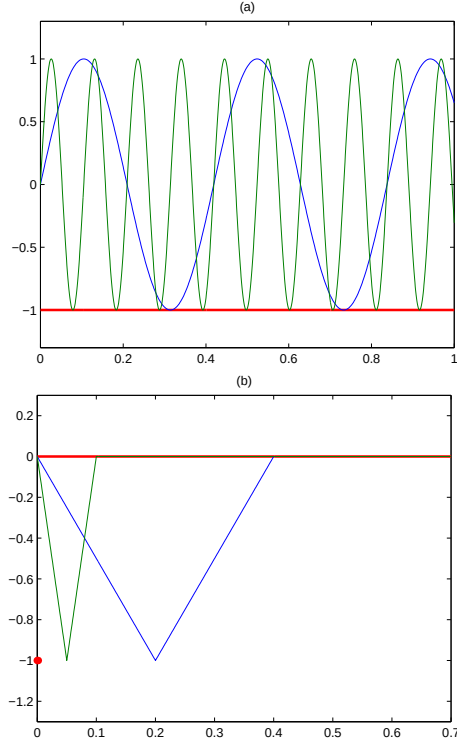


Figure 5. Examples of Γ -convergence: (a) $\sin nx$, (b) f_n .

- If $\Gamma\text{-lim } F_i = F$, f_i minimizes F_i and $f_i \rightarrow f$, f minimizes F . More than that, it is enough to have $F_i(f_i) \leq \inf F_i + \varepsilon_i$ for some $\varepsilon_i \downarrow 0$.
- Suppose that $\Gamma\text{-lim } F_i = F$ and that there is a compact set $K \subset X$ such that $\forall i \inf_X F_i = \inf_K F_i$. Then there exists $\min_X F = \lim_i \inf_X F_i$. Moreover, if f_i is a minimizer of F_i and $f_i \rightarrow f$, then f is a minimizer of F .

3.2.2. The approximations

We come back to the task of approximating the Mumford-Shah functional by a nicer functional. However, we can not approximate $F(u)$ with functionals of the usual local integral form $F_\varepsilon(u) = \int_\Omega f_\varepsilon(\nabla u, u)$ for $u \in W^{1,1}$ (see [19, p. 56]). One of the possibilities to overcome this is to introduce a second auxiliary variable, which was done in [6, 7].

The approximation proposed in [7] is

$$F_\varepsilon(u, v) = \int_\Omega \left[v^2 |\nabla u|^2 + \beta \left(\frac{(v-1)^2}{4\varepsilon} + \varepsilon |\nabla v|^2 \right) + \alpha |u - w|^2 \right] dx. \quad (3)$$

The meaning of v in this functional is clear—it approximates $1 - \chi_{S_u}$, being close to 0 when $|\nabla u|$ is large and 1 otherwise. This functional is elliptic and is relatively easy to minimize numerically, for an implementation see [61].

A finite-element discretization was proposed in [10], with a proof that the discretized functionals also Γ -converge to $F(u)$ if the mesh-size is $o(\varepsilon)$. This discretization was improved and implemented in [17].

Other possible approximations of F can be found in [19]. Some of these approximating functionals are easier to minimize using finite elements and not finite differences. It should be taken into account, though, that the geometry of the mesh is important and in general can introduce anisotropy to the limiting functional [59, 18]. On the other hand, finite elements implementations with adaptive mesh based on non-local approximations have less parameters to take care of compared to finite differences.

A different approach is taken in [77], which represents S_u as a curve using level sets, and the curve evolves together with the image itself to minimize the Mumford-Shah functional by steepest descent.

The piecewise-constant approximation $\bar{F}$, Eq. (2), has also been a subject of numerical implementations. A region-merging algorithm was proposed in [47] (also described in [55]); it creates a scale-space-like pyramidal structure with respect to the parameter β . Another implementation using level-set methods to represent S_u is described in [25] and then extended in [26]. In these works, however, the number of different values u can assume must be prescribed *a priori*.

3.3. RELATED FORMULATIONS

In [12] Blake and Zisserman describe models of piecewise-continuous images. One of these models—the weak membrane—is the Mumford-Shah functional. The discretization of the functional they suggest is

$$\sum W(|f_{i+1,j} - f_{i,j}|) + W(|f_{i,j+1} - f_{i,j}|) + |f_{i,j} - g_{i,j}|^2$$

where W is given by

$$W = \min\{\mu^2 x^2, \nu\}, \quad \mu, \nu \text{ parameters.}$$

They also suggest an algorithm (the graduated non-convexity) to minimize the discretized functional. In [20, 21] the relationship between

the Mumford-Shah and Blake-Zisserman formulations is clarified: the Blake-Zisserman discrete formulation Γ -converges to a functional close to the Mumford-Shah functional. However, the distinction is important: in the limiting case of Blake-Zisserman functional the edge length is replaced by the sum of its projections on the axes. Another similar discrete model, due to Geman and McClure, also Γ -converges to the same functional [62].

In [68] the functional $\int_{\Omega} (f - g)^2 + \lambda(|\nabla f|)$ is proposed, where $\lambda(t)$ is quadratic for $t < c_{\rho}$ and linear for $t \geq c_{\rho}$ (c_{ρ} is a parameter determining the coarseness and amount of edges to be detected).

The authors of [27] propose a discrete segmentation model, based upon certain conditions that guarantee that the model is edge-preserving. They then represent this model as the infimum of quadratic functionals by introducing an auxiliary variable that “plays the role of a discontinuity marker” (half-quadratic regularization). In a subsequent paper [76] they add a smoothing constraint for this variable, thus arriving at a functional almost identical to Eq. (3).

In [65] the authors propose the functional

$$J_{\varepsilon} = \int_{\Omega} |f - g|^2 + \varepsilon \lambda^2 \int_{\Omega} \varphi(|\nabla f|) + \frac{\eta^2}{\varepsilon} \int_{\Omega} W(f),$$

where $\varphi(|\nabla f|)$ is an edge-preserving smoothing term (similar to those discussed in Section 4.4.2), and W is a multi-well potential, for segmentation of images consisting of patches belonging to a few classes (many satellite images belong to this category). The number and the characteristics of these classes (mean brightness and standard deviation) are known beforehand. As in the above two works, the authors introduce an auxiliary variable based on the half-quadratic regularization, and arrive at

$$J_{\varepsilon}^*(f, b) = \int_{\Omega} |f - g|^2 + \varepsilon \lambda^2 \int_{\Omega} [b|\nabla f| + \psi(b)] + \frac{\eta^2}{\varepsilon} \int_{\Omega} W(f),$$

which they minimize numerically. The authors mention some Γ -convergence results to justify their model, but do not prove the Γ -convergence of J_{ε} or of J_{ε}^* . In [64] this model is extended to the vectorial case.

A term penalizing for curvature ($\int_K \kappa^2$) is added to the Mumford-Shah functional in [52]. The authors show Γ -convergence and give numerical examples.

The author of [70] modified the Ambrosio-Tortorelli construction to be

$$G_{\varepsilon} = \int_{\Omega} \left[v^2 |\nabla u| + \alpha |u - w| + \beta \left(\varepsilon |\nabla v|^2 + \frac{(1 - v)^2}{4\varepsilon} \right) \right]$$

and conjectured that it approximates in the sense of Γ -convergence the functional

$$\int_{\Omega \setminus K} (|\nabla u| + \alpha|u - w|) dx + \beta \int_K \frac{|u^+ - u^-|}{1 + |u^+ - u^-|} d\mathcal{H}^1.$$

This result is proved in [1].

4. Generalizing Mumford-Shah functional to color

4.1. THE STRAIGHTFORWARD GENERALIZATION

The most obvious way to generalize the Mumford-Shah functional to color images $u : \Omega \rightarrow \mathbb{R}^3$ is to use

$$F(u) = \int_{\Omega} (|\nabla u|^2 + \alpha|u - w|^2) + \beta \mathcal{H}^{n-1}(S_u). \quad (4)$$

In this case the only coupling between the channels is through the common jump set S_u . The approximation results from Section 3.2 translate to this case without change (as noted in [7]) and we can use the elliptic approximation

$$F_{\varepsilon}(u, v) = \int_{\Omega} \left[v^2 |\nabla u|^2 + \beta \left(\frac{(v-1)^2}{4\varepsilon} + \varepsilon |\nabla v|^2 \right) + \alpha|u - w|^2 \right] dx, \quad (5)$$

to find minimizers of $F(u)$. We minimize F_{ε} by steepest descent,

$$\begin{aligned} u_t &= -C_u [2\alpha(u - w) - \operatorname{div}(2v^2 \nabla u)] \\ &= -2C_u [\alpha(u - w) - v^2 \Delta u - 2v \langle \nabla v, \nabla u \rangle], \\ v_t &= -C_v \left[2v |\nabla u|^2 + \beta \left(\frac{v-1}{2\varepsilon} - 2\varepsilon \Delta v \right) \right]. \end{aligned}$$

A result of numerical minimization is shown in Figure 6. The original image is shown in Figure 7.

4.2. THE GEOMETRICAL GENERALIZATIONS

We want to generalize the Mumford-Shah functional

$$\int_{\Omega} (|\nabla u|^2 + \alpha|u - w|^2) + \beta \mathcal{H}^{n-1}(S_u)$$

to color images, using the “image as a manifold” interpretation, while the length term $\mathcal{H}^{n-1}(S_u)$ remains the same.

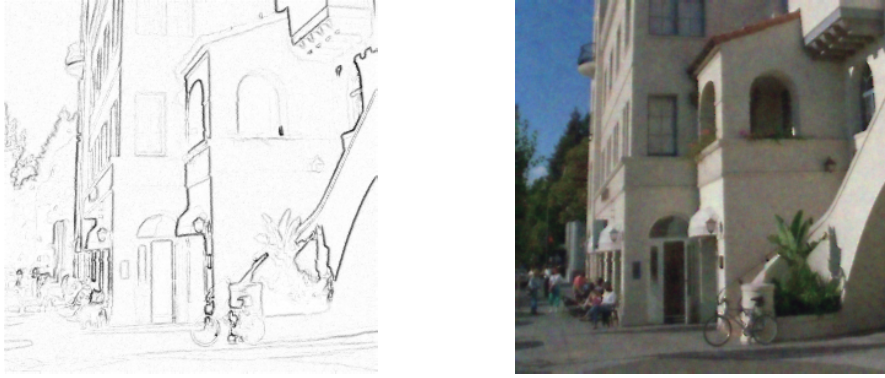


Figure 6. Results of numerical minimization of F , with $\varepsilon = 0.04$, $\alpha = 0.6$, $\beta = 0.015$.



Figure 7. The original image.

The fidelity term that is most consistent with the geometric approach would be the Hausdorff distance between the two surfaces, or at least $\int_{\Omega} d(u(x), w(x))$, where $d(\cdot, \cdot)$ is the geodesic distance in the feature space, as in [15]. Yet, both these approaches seem computationally intractable. The suggestion of $\int_{\Omega} \|u - w\|_h^2$, made in [45], (here h_{ij} is the metric on the feature space, and $\|\cdot\|_h$ is the corresponding norm on the tangent space) is easy to implement, but lacks mathematical

validity: $u - w$ is not in the tangent space. We will use the simplest reasonable alternative, $\int_{\Omega} |u - w|^2$.

The smoothing term is the area $\int_{\Omega} \sqrt{\det g}$ or the energy $\int_{\Omega} \det g$, where g is the metric induced on the manifold by h —the metric on the feature space. In the case where

$$h = \begin{bmatrix} \gamma & 0 & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

is a Euclidean metric on $\Omega \times [0, 1]^3$ and

$$U(x, y) = (x, y, R, G, B) = (x, y, u^1(x, y), u^2(x, y), u^3(x, y))$$

is the embedding, we get

$$\begin{aligned} \det g &= \det(dU^* \circ h \circ dU) \\ &= \gamma^2 + \gamma \sum_i |\nabla u^i|^2 + \sum_{i,j} |\nabla u^i \times \nabla u^j|^2 \\ &= \gamma^2 + \gamma(|u_x|^2 + |u_y|^2) + |u_x \times u_y|^2 \\ &= \gamma^2 + \gamma|\nabla u|^2 + |u_x \times u_y|^2. \end{aligned}$$

Thus, we have two models:

$$\begin{aligned} F^1(u) &= \int_{\Omega} \sqrt{\gamma^2 + \gamma|\nabla u|^2 + |u_x \times u_y|^2} + \alpha \int_{\Omega} |u - w|^2 + \beta \mathcal{H}^{n-1}(S_u), \\ F^2(u) &= \int_{\Omega} (\gamma|\nabla u|^2 + |u_x \times u_y|^2) + \alpha \int_{\Omega} |u - w|^2 + \beta \mathcal{H}^{n-1}(S_u). \end{aligned}$$

Note that γ^2 was dropped in the second functional, since in this case it merely adds a constant to the functional.

However, the theory of functionals on SBV or GSBV seems to be unable to deal with these models at the moment. It is necessary to establish lower semicontinuity of the functionals, both to ensure existence of minimizers, and as an important component in the Γ -convergence proofs. Though, theorems on lower semicontinuity of functionals on SBV exist only for isotropic functionals (depending only on $|\nabla u|$ and not on ∇u itself), or at least functionals with constant rate of growth, i.e.

$$c|\nabla u|^r \leq f(\nabla u) \leq C(1 + |\nabla u|)^r \text{ for some } C > c > 0 \text{ and } r > 1.$$

The term $|u_x \times u_y|^2$ is of order $|\nabla u|^4$, yet we can not bound it from below by $c|\nabla u|^4$ for some $c > 0$, therefore we can not use these theorems.

The role of the term $|u_x \times u_y|^2$ is explored in [44]. If we assume the Lambertian light reflection model, then $u(x, y) = \langle n(x, y), l \rangle \rho(x, y)$, where $n(x, y)$ is the unit normal to the surface, l is the light source direction, and $\rho(x, y)$ captures the characteristics of the material. Assuming that for any given object $\rho(x, y) = \rho = \text{const}$ we have $u(x, y) = \langle n(x, y), l \rangle \rho$, hence $\text{Im } u \subset \text{span}\{\rho\}$ and $\text{rank } du \leq 1$. This is equivalent to $u_x \times u_y = 0$.

Thus, the term $|u_x \times u_y|^2$ in the edge indicator enforces the Lambertian model on every smooth surface patch. It also means that taking rather small γ makes sense, since we expect $|u_x \times u_y|^2$ to be (almost) 0, and $|\nabla u|^2$ to be just small.

4.3. THE PROPOSED GENERALIZATIONS

A generalization of the Mumford-Shah functional proposed here is an attempt to combine the nice smoothing-segmenting features of the geometric model with the existing Γ -convergence results for the elliptic approximation of the original Mumford-Shah functional. We pay for that by the loss of some of the geometric intuition behind the manifold interpretation. First, we replace $|u_x \times u_y|^2$ by $|u_x \times u_y|$ in F^1 and F^2 :

$$\begin{aligned} G^1(u) &= \int_{\Omega} (\gamma |\nabla u|^2 + |u_x \times u_y|) + \alpha \int_{\Omega} |u - w|^2 + \beta \mathcal{H}^{n-1}(S_u), \\ G^2(u) &= \int_{\Omega} \sqrt{\gamma^2 + \gamma |\nabla u|^2 + |u_x \times u_y|} + \alpha \int_{\Omega} |u - w|^2 + \beta \mathcal{H}^{n-1}(S_u). \end{aligned}$$

Note that $|u_x \times u_y|$ enforces the Lambertian model, just as $|u_x \times u_y|^2$.

The new functional G^2 seems to violate another important requirement, necessary for lower semicontinuity with respect to L^1 convergence: being quasiconvex (see the definition and discussion in Section 4.3.1). Besides, since the smoothing term is of linear growth, approximation similar to those in Section 3.2 will converge to a functional with more interaction between the area and the length terms, and depending on the Cantor part of Du . We thus propose the functional

$$G^3(u) = \int_{\Omega} \sqrt{\gamma + |\nabla u|^2} + \alpha \int_{\Omega} |u - w|^2 + \beta \int_{S_u} \frac{|u^+ - u^-|}{1 + |u^+ - u^-|} d\mathcal{H}^{n-1} + |D^c u|(\Omega).$$

Although lacking the term $|u_x \times u_y|$, it nevertheless enforces quite strong coupling between the different channels, definitely stronger than $F(u)$, Eq. (4) (in $F(u)$ the coupling is only through the common edges, and in G^3 the smoothing term introduces additional coupling).

4.3.1. Analysis and implementation of G^1

The elliptic approximation for G^1 is provided in [40]. Actually, a much more general result is proven, but what we need is the following

Theorem. Let $f : \mathbb{R}^{n \times m} \rightarrow [0, +\infty)$ be quasiconvex and satisfy $c_1|z|^p \leq f(z) \leq c_2|z|^p$ with $p > 1$, c_1, c_2 positive constants. Let the functionals $G_\varepsilon(u, v)$ be defined by

$$G_\varepsilon(u, v) = \int_{\Omega} \left[(|v|^2 + \eta_\varepsilon) f(\nabla u) + \frac{\varepsilon^{p-1}}{p} |\nabla v|^p + \frac{(v-1)^p}{a\varepsilon p'} \right] dx,$$

if $u, v \in W^{1,p}$, $0 \leq v \leq 1$, and $G_\varepsilon(u, v) = +\infty$ otherwise. Here $p' = \frac{p}{p-1}$, $a = (2 \int_0^1 (1-s)^{p-1} ds)^{p'}$, $\eta_\varepsilon = o(\varepsilon^{p-1})$, $\eta_\varepsilon > 0$. Then $G_\varepsilon(u, v)$ Γ -converge to $G(u, v)$ defined by

$$G(u, v) = \int_{\Omega} f(\nabla u) dx + \mathcal{H}^{n-1}(S(u))$$

if $u \in \text{GSBV}(\Omega, \mathbb{R}^m)$ and $v = 1$ a.e. and $G(u, v) = +\infty$ otherwise. The Γ -convergence is with respect to the convergence in measure.

We will now check that all the conditions of the theorem are satisfied.

A function $f : \mathbb{R}^{n \times m} \rightarrow [0, +\infty)$ is said to be *quasiconvex* if for every $a \in \mathbb{R}^{n \times m}$ and for every $\varphi \in W_0^{1,\infty}(D, \mathbb{R}^m)$ and for every bounded domain $D \subset \mathbb{R}^n$ we have

$$f(A) \leq \int_D f(A + \nabla \varphi(x)) dx \Big/ \int_D dx.$$

This condition is very difficult to verify. We will use another property, polyconvexity, which is easier to check and which implies quasiconvexity. A function f is called *polyconvex* if there exists a convex function g such that

$$f(A) = g(A, \text{adj}_2 A, \dots, \text{adj}_{\min\{m,n\}} A).$$

Here $\text{adj}_k A$ is the matrix of all $k \times k$ minors of the matrix A . The book [30] can be used as a reference on this topic. In our case

$$f = \gamma |\nabla u|^2 + |u_x \times u_y| = \gamma |\nabla u|^2 + |\text{adj}_2 \nabla u|,$$

so f is polyconvex and thus quasiconvex.

The growth estimate holds with $p = p' = 2$ since

$$\begin{aligned} \gamma |\nabla u|^2 &\leq \gamma |\nabla u|^2 + |u_x \times u_y| = \gamma |\nabla u|^2 + |u_x| |u_y| |\sin \theta| \\ &\leq \gamma |\nabla u|^2 + \frac{|\sin \theta|}{2} |\nabla u|^2 \leq \left(\gamma + \frac{1}{2} \right) |\nabla u|^2. \end{aligned}$$

Thus $a = 1$, and we have the approximation

$$G_\varepsilon(u, v) = \int_{\Omega} \left[(v^2 + \eta_\varepsilon) (\gamma |\nabla u|^2 + |u_x \times u_y|) + \frac{\varepsilon}{2} |\nabla v|^2 + \frac{(v-1)^2}{2\varepsilon} + \alpha |u - w|^2 \right] dx.$$

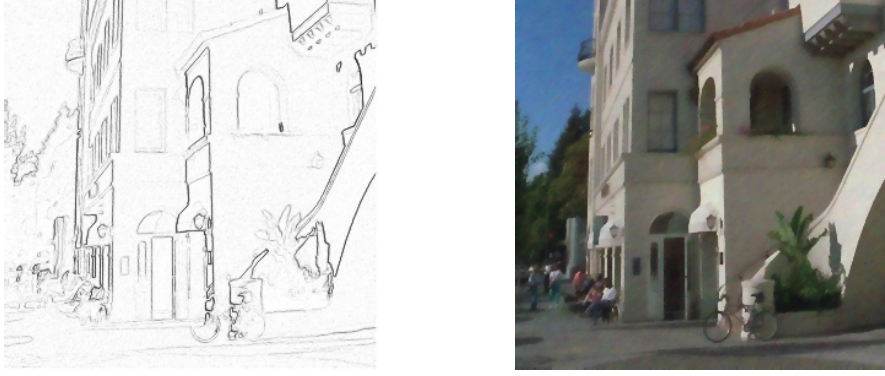


Figure 8. Minimization of G^1 with $\varepsilon = 0.3$, $\alpha = 0.05$, $\beta = 0.0025$, $\gamma = 0.01$.

We will neglect $\eta_\varepsilon = o(\varepsilon)$ (as in [61, 74] and other works). As noted in [69], the function of η_ε is to ensure regularity of minimizers and not the convergence of the functionals. We will also change slightly the part that approximates the edges: $\frac{\varepsilon}{2}|\nabla v|^2 + \frac{(v-1)^2}{2\varepsilon}$; we will multiply the first term by 2 and divide the second by the same factor, to stay closer to the original approximation (5). We also have to multiply this part by β . These changes do not affect the Γ -convergence proof, and the approximation of G^1 is

$$G_\varepsilon^1(u, v) = \int_{\Omega} \left[v^2(\gamma|\nabla u|^2 + |u_x \times u_y|) + \beta \left(\varepsilon|\nabla v|^2 + \frac{(v-1)^2}{4\varepsilon} \right) + \alpha|u-w|^2 \right] dx.$$

The steepest descent equations are

$$\begin{aligned} u_t &= -C_u \left[2\alpha(u-w) - \operatorname{div} \left\{ v^2 \left(2\gamma u_x + \frac{u_y \times (u_x \times u_y)}{|u_x \times u_y|}, 2\gamma u_y - \frac{u_x \times (u_x \times u_y)}{|u_x \times u_y|} \right) \right\} \right] \\ v_t &= -C_v \left[2v(\gamma|\nabla u|^2 + |u_x \times u_y|) + \beta \left(\frac{v-1}{2\varepsilon} - 2\varepsilon\Delta v \right) \right]. \end{aligned}$$

Results of a numerical minimization of G^1 are shown in Figure 8.

4.3.2. Analysis and implementation of G^3

A functional similar to the Mumford-Shah functional, but with linear growth in the gradient is examined in [1], and it is proved in particular

that $\Gamma\text{-lim } G_\varepsilon = G$ (with respect to L^1 convergence), where

$$\begin{aligned} G_\varepsilon(u, v) &= \int_\Omega \left[v^2 f(|\nabla u|) + \beta \left(\varepsilon |\nabla v|^2 + \frac{(1-v)^2}{4\varepsilon} \right) \right] \\ &\quad \text{if } u, v \in H^1(\Omega) \text{ and } 0 \leq v \leq 1 \text{ a.e., and } +\infty \text{ otherwise,} \\ G(u, v) &= \int_\Omega f(|\nabla u|) + \beta \int_{S_u} \frac{|u^+ - u^-|}{1 + |u^+ - u^-|} d\mathcal{H}^1 + |D^c u|(\Omega) \\ &\quad \text{if } u \in \text{GBV}(\Omega) \text{ and } v = 1 \text{ a.e., and } +\infty \text{ otherwise,} \end{aligned}$$

and $f : [0, +\infty) \rightarrow [0, +\infty)$ is convex, increasing, and $\lim_{z \rightarrow \infty} f(z)/z = 1$. With the aim of generalizing this result to color images we define $f(z) = \sqrt{\gamma + z^2}$,

$$\begin{aligned} G^3(u) &= \int_\Omega \sqrt{\gamma + |\nabla u|^2} + \alpha \int_\Omega |u - w|^2 + \beta \int_{S_u} \frac{|u^+ - u^-|}{1 + |u^+ - u^-|} d\mathcal{H}^{n-1} + |D^c u|(\Omega), \\ G_\varepsilon^3(u, v) &= \int_\Omega \left[v^2 \sqrt{\gamma + |\nabla u|^2} + \alpha |u - w|^2 + \beta \left(\varepsilon |\nabla v|^2 + \frac{(1-v)^2}{4\varepsilon} \right) \right], \end{aligned}$$

with domains as above.

Upon inspection of the proofs in [1], it seems that everything remains valid for the vectorial case, except one part, that establishes the lower inequality for the one-dimensional case ($n = 1$) in a small neighborhood of a jump point. We will now provide a “replacement” for this part (the second part of Proposition 4.3 in [1], beginning with Eq. (4.4)). We will need the Jensen’s inequality in integral form for a function of several variables (see e.g. [41]): if $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$ is convex, and $(u^1, \dots, u^m) : \mathbb{R} \rightarrow \mathbb{R}^m$ are integrable, then

$$\varphi\left(\frac{1}{b-a} \int_a^b u^1, \dots, \frac{1}{b-a} \int_a^b u^m\right) \leq \frac{1}{b-a} \int_a^b \varphi(u^1, \dots, u^m). \quad (6)$$

Proof. We consider $G_{\varepsilon_j}(u_j, v_j, (t - \eta, t + \eta))$. For any $\delta > 0$ we can find $x_1, x_2 \in (t - \eta, t + \eta)$ such that

$$\begin{aligned} \lim |u_j(x_1) - u_j(x_2)| &> \operatorname{ess\,sup}_{\tau_1, \tau_2 \in (t-\eta, t+\eta)} |u(\tau_1) - u(\tau_2)| - \delta, \\ \lim v_j(x_1) &= \lim v_j(x_2) = 1. \end{aligned}$$

Take $\bar{x}_j \in [x_1, x_2]$ such that $v_j(\bar{x}_j) = \inf_{[x_1, x_2]} v_j < 1$. Then we obtain the following estimate:

$$\begin{aligned}
G_{\varepsilon_j}(u_j, v_j, (t - \eta, t + \eta)) &\geq G_{\varepsilon_j}(u_j, v_j, (x_1, x_2)) \\
&= \int_{x_1}^{x_2} v^2 f(|\nabla u_j|) + \beta \left(\varepsilon |\nabla v_j|^2 + \frac{(1 - v_j)^2}{4\varepsilon} \right) \\
&\geq v_j(\bar{x}_j)^2 \int_{x_1}^{x_2} f(|\nabla u_j|) + \beta \int_{x_1}^{x_2} \left(\varepsilon |\nabla v_j|^2 + \frac{(1 - v_j)^2}{4\varepsilon} \right) \\
&\geq v_j(\bar{x}_j)^2 |x_2 - x_1| f\left(\frac{|u(x_2) - u(x_1)|}{|x_2 - x_1|}\right) + \beta \int_{x_1}^{x_2} (1 - v_j) |v'_j| dt \\
&\geq v_j(\bar{x}_j)^2 |x_2 - x_1| \frac{|u(x_2) - u(x_1)|}{|x_2 - x_1|} + \beta \int_{v_j(\bar{x}_j)}^{v_j(x_1)} (1 - s) ds + \beta \int_{v_j(\bar{x}_j)}^{v_j(x_2)} (1 - s) ds \\
&\geq \inf_{\tau \in [0, 1]} \left\{ \tau^2 |u(x_2) - u(x_1)| + \beta \int_{\tau}^{v_j(x_1)} (1 - s) ds + \beta \int_{\tau}^{v_j(x_2)} (1 - s) ds \right\}.
\end{aligned}$$

We have used Eq. (6) for the function $\varphi(z) = f(|z|)$ to establish

$$\begin{aligned}
\int_{x_1}^{x_2} f(|\nabla u_j|) &= \int_{x_1}^{x_2} \varphi(\nabla u_j) \geq |x_1 - x_2| \varphi\left(\frac{\int_{x_1}^{x_2} \frac{du^1}{dx}}{x_2 - x_1}, \dots, \frac{\int_{x_1}^{x_2} \frac{du^m}{dx}}{x_2 - x_1}\right) \\
&= |x_1 - x_2| \varphi\left(\frac{u^1(x_2) - u^1(x_1)}{x_2 - x_1}, \dots, \frac{u^m(x_2) - u^m(x_1)}{x_2 - x_1}\right) \\
&= |x_1 - x_2| f\left(\frac{1}{|x_2 - x_1|} |u(x_2) - u(x_1)|\right).
\end{aligned}$$

We have also used the inequality $\mu^2 + \nu^2 \geq 2\mu\nu$ with $\mu = \sqrt{\varepsilon} |\nabla v_j|$ and $\nu = \frac{1-v_j}{2\sqrt{\varepsilon}}$, and in the next line—the fact that in our case $f(z) > z$.³

Letting $j \rightarrow \infty$, we get

$$\begin{aligned}
&\liminf G_{\varepsilon_j}(u_j, v_j, (t - \eta, t + \eta)) \\
&\geq \inf_{\tau \in [0, 1]} \left\{ \tau^2 \left| \operatorname{ess\,sup}_{\tau_1, \tau_2 \in (t - \eta, t + \eta)} |u(\tau_1) - u(\tau_2)| - \delta \right| + 2\beta \int_{\tau}^1 (1 - s) ds \right\}.
\end{aligned}$$

From here we can proceed as in [1].

Recently, [2] provided a more general result for linear-growth functionals, which also implies our convergence result for G^3 . The proofs in [2] are much more technical than in [1].

³ The condition $f(z) > z$ is not necessary to establish this result, but makes it a bit easier. We could alternatively use $\lim_{z \rightarrow \infty} f(z)/z = 1$, infer that $f(z) \geq z(1 - \theta(z))$, $\lim_{z \rightarrow \infty} \theta(z) = 0$, and wait until we let $\eta \rightarrow 0$ (and thus $x_2 \rightarrow x_1$) for $\theta(z)$ to disappear.

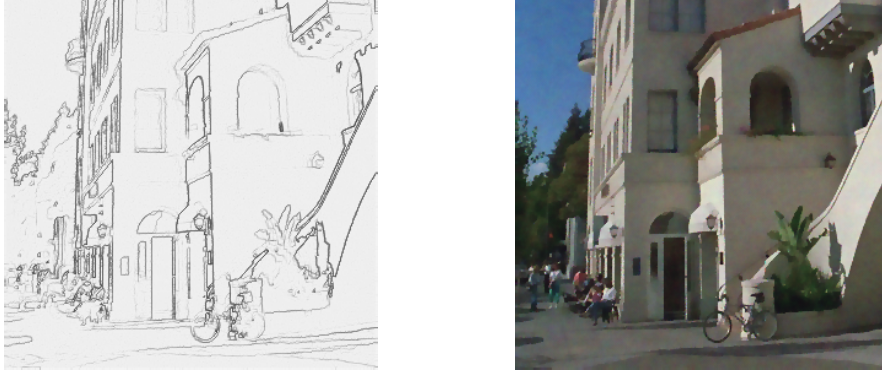


Figure 9. Minimization of G^3 with $\varepsilon = 0.3$, $\alpha = 0.05$, $\beta = 0.0025$, $\gamma = 0.01$.

The steepest descent equations for G_ε^3 are

$$\begin{aligned} u_t &= -C_u \left[2\alpha(u - w) - \operatorname{div} \left(\frac{v^2(u_x, u_y)}{\sqrt{\gamma + |\nabla u|^2}} \right) \right], \\ v_t &= -C_v \left[2v\sqrt{\gamma + |\nabla u|^2} + \beta \left(\frac{v-1}{2\varepsilon} - 2\varepsilon\Delta v \right) \right]. \end{aligned}$$

Results of a numerical minimization of G^3 are shown in Figure 9.

4.4. GENERAL CONSIDERATIONS

4.4.1. Details of numerics

All the implementations demonstrated here are very straightforward. The steepest descent equations are discretized, and Euler method (forward derivative in time) is used with timestep small enough to ensure stability. Neumann (natural) boundary conditions are used. Typical running time for the implementations shown here were 1 to 5 minutes (depending on the functional and the parameters) in MATLAB on a Pentium-500.

The only non-trivial detail is that the terms of the form $\operatorname{div}(C\nabla u)$ should be discretized with forward derivatives for the gradient and backwards derivatives for the divergence, or vice versa, to provide tight numerical support.

Another aspect of the numerics is the selection of ε . Consider the simplest one-dimensional case

$$F_\varepsilon(u, v) = \int_{\Omega} \left[v^2 |\nabla u|^2 + \beta \left(\frac{(v-1)^2}{4\varepsilon} + \varepsilon |\nabla v|^2 \right) + \alpha |u - w|^2 \right] dx.$$

On one hand, we want ε to be small to have the best possible approximation of the original functional; on the other hand, taking ε too small

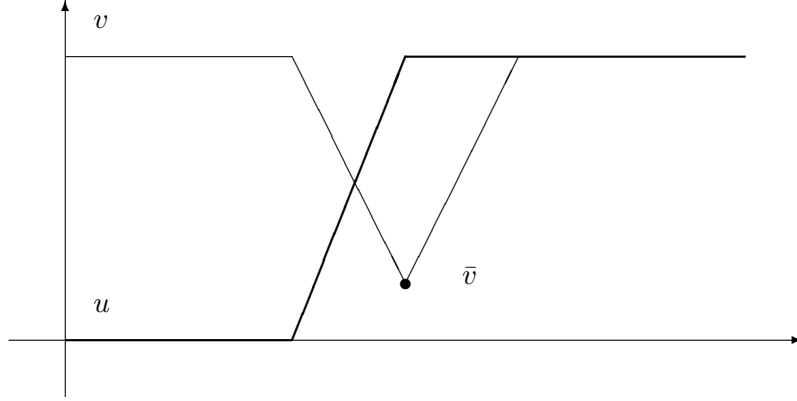


Figure 10. The simplest one-dimensional case: a jump.

will give too much weight to the $\frac{(v-1)^2}{4\varepsilon}$ term, thus preventing edge detection.

Let us consider a case when w is a “pure jump” from 0 to 1, and α is high enough to force a jump in u . In this case, we expect the resulting u and v to be as in Figure 10, u being (almost) equal to w , and v being not equal to 1 at one point (the situation will be as shown, if all derivatives are backwards).

Let $\bar{v}$ be the value of v at the jump point $0 \leq \bar{v} < 1$. If we discretize the functional $F_\varepsilon(u, v)$ (using backward derivatives) the only significantly different from zero terms will be

$$f(\bar{v}) = \frac{\bar{v}^2}{h^2} + \beta \left[\frac{(1 - \bar{v})^2}{4\varepsilon} + 2\varepsilon \left(\frac{1 - \bar{v}}{h} \right)^2 \right].$$

The minimal value of $f(\bar{v})$ is at $\bar{v} = \frac{B}{A+B}$, where $A = 1/h^2$ and $B = \beta[\frac{1}{4\varepsilon} + \frac{2\varepsilon}{h^2}]$. If we want the jump to be detected, that is, to have $\bar{v} \ll 1$, we must have $B \ll A$. This means that $\beta h^2 \ll \varepsilon$ and $\varepsilon\beta \ll 1$ must hold. These are indeed true for the results shown in Section 4.1 and Section 4.3.1 (in our implementations we always took $h = 1$).

4.4.2. On one common theme in image processing

Many modern algorithms for image restoration have one detail in common. It is introduced by different authors, based on different assumptions or observations, and in some cases without noticing that the same idea has guided other researchers. It is the desire to have a smoothing term that will be the $W^{1,1}$ norm near the edges, and the $W^{1,2}$ norm far from them.

Many different techniques were employed to achieve this purpose. In [81] the authors introduce an edge indicator function $f(|\nabla u|)$ and set

$$f'(x) = \frac{\sqrt{x^2 + b^2} - \sqrt{(x - a)^2 + b^2} - b + \sqrt{a^2 + b^2}}{2a}.$$

For small b (the values $a = 5$, $b = 0.01$ were used) $f(x)$ is very close to a quadratic function on $[0, a]$ and to a linear function on $[a, +\infty)$.

It is pointed out in [11] that the $f(x)$ is known in statistics as the Huber's minimax estimator, designed to be an error measure not too sensitive to outliers.

A similar function appears in [22], in an approximation to the total variation. To overcome staircasing (see Section 2.1.2) the same article also suggests minimizing

$$\frac{1}{2\varepsilon} \int_{|\nabla u| < \varepsilon} |\nabla u|^2 + \int_{|\nabla u| \geq \varepsilon} |\nabla u| - \frac{\varepsilon}{2},$$

which is actually the same approximation (compare Eq. (16) and Eq. (31) in [22]), but here ε is taken to be rather large.

The model proposed in [14] is

$$\int_{\Omega} |\nabla u|^{p(|\nabla u|)},$$

where $p(t)$ is a smooth monotone function, descending from 2 to 1 as t grows.

In [68] we see again a smoothing term of the form $\int_{\Omega} \lambda(|\nabla f|)$, where $\lambda(t)$ is quadratic for $t < c_{\rho}$ and linear for $t \geq c_{\rho}$. Here c_{ρ} is a parameter determining the coarseness and amount of edges to be detected, and is to be set by the user.

Finally, we studied in this work edge indicators like $\sqrt{\gamma^2 + \gamma|\nabla u|^2 + |u_x \times u_y|^2}$ (after [72]) and $\sqrt{\gamma + |\nabla u|^2}$. It is noted in [72, 73] that for $|\nabla u| \ll \sqrt{\gamma}$ we have $\sqrt{\gamma + |\nabla u|^2} \approx \sqrt{\gamma} + \frac{|\nabla u|^2}{2\sqrt{\gamma}}$, and for $|\nabla u| \gg \sqrt{\gamma}$ we have $\sqrt{\gamma + |\nabla u|^2} \approx |\nabla u|$. Thus, we again have a quadratic function in the smooth regions, and a linear one near the edges, with $\sqrt{\gamma}$ being the threshold. In [73] the parameter γ is changed locally, depending on the image.

To the best of our knowledge, there were no attempts to unify these approaches into one mathematically solid framework. Hopefully, it will become a topic of future research.

5. Summary and future research

5.1. SUMMARY

In this work we suggested, analyzed and implemented some possible generalizations of the Mumford-Shah functional to the color images, based on a geometric model of images as manifolds.

It seems that if we want to have a Γ -convergence of a numerical approximation to the suggested models, the models must be simplified, at least at the present state of the theory of the Γ -convergence for the free-discontinuity problems.

Some of the numerical results are quite satisfactory. It is easy to see how different models accentuate different features in edge detection and restoration. It seems that none of the suggested models is the best, but rather that a model should be selected depending on the task at hand.

5.2. UNANSWERED QUESTIONS AND UNEXPLORED POSSIBILITIES

In spite of all theoretical difficulties we encountered, we can write “approximations” modeled after Eq. (3) for any reasonable edge indicator function, and minimize them numerically. Since the intuition behind the approximation is clear, we can give a fairly good (albeit not rigorous) qualitative prediction of the result, the way it was done in [45, 76]. We present in Figure 11 as an example a result of numerical minimization of

$$F_\varepsilon^2(u, v) = \int_\Omega \left[v^2(\gamma|\nabla u|^2 + |u_x \times u_y|^2) + \beta \left(\varepsilon|\nabla v|^2 + \frac{(v-1)^2}{4\varepsilon} \right) + \alpha|u-w|^2 \right] dx.$$

Since the results are good, it is reasonable to expect that we can prove convergence for F_ε^2 . At the moment, we can only show that F^2 is lower semicontinuous. Also, we do not know, if anything interesting can be shown about F^1 and G^2 .

Some other questions and possible improvements:

- We do not have convergence results for the numerical method we currently use. For some other method of minimization it might be possible to prove convergence and give an error estimate.
- The current numerical method is also rather slow and definitely can be improved. One simple possibility is to vary the time step in a kind of a secant method.
- We could have had “edge focusing” like in [61].

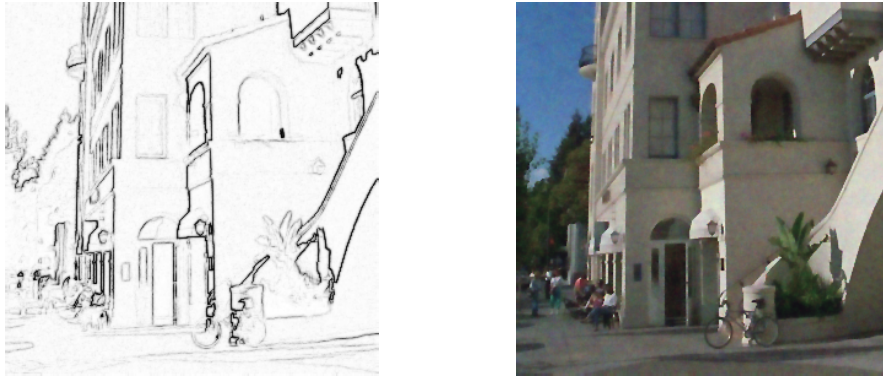


Figure 11. Minimization of F^2 with $\varepsilon = 0.3$, $\alpha = 0.1$, $\beta = 0.01$, $\gamma = 0.2$.

- Setting the problem in other color spaces—HSV (see [50]), YCC, CB, CIELAB—may improve the results.

References

1. Alicandro, R., A. Braides, and J. Shah: 1999, ‘Free-discontinuity problems via functionals involving the L^1 -norm of the gradient and their approximation’. *Interfaces and Free Boundaries* **1**(1), 17–37.
2. Alicandro, R. and M. Focardi: 2001, ‘Variational Approximation of Free-Discontinuity Energies with linear growth’. Technical report, CVGMT, Pisa.
3. Ambrosio, L., L. Faina, and R. March: 2001, ‘Variational approximation of a second order free discontinuity problem in computer vision’. *SIAM J. Math. Anal.* **32**(6), 1171–1197.
4. Ambrosio, L., N. Fusco, and D. Pallara: 1999, ‘Higher regularity of solutions of free discontinuity problems’. *Differential Integral Equations* **12**(4), 499–520.
5. Ambrosio, L., N. Fusco, and D. Pallara: 2000, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs. Oxford University Press.
6. Ambrosio, L. and V. M. Tortorelli: 1990, ‘Approximation of functionals depending on jumps by elliptic functionals via Γ -convergence’. *Comm. Pure Appl. Math.* **43**(8), 999–1036.
7. Ambrosio, L. and V. M. Tortorelli: 1992, ‘On the approximation of free discontinuity problems’. *Boll. Un. Mat. Ital. B (7)* **6**(1), 105–123.
8. Bakushinsky, A. and A. Goncharky: 1994, *Ill-posed problems: theory and applications*. Dordrecht: Kluwer Academic Publishers.
9. Ballester, C. and M. González: 1996, ‘Texture segmentation by variational methods’. In: *ICAOS '96. 12th International Conference on Analysis and Optimization of Systems. Images, Wavelets and PDEs*. pp. 187–193.
10. Bellettini, G. and A. Coscia: 1994, ‘Discrete approximation of a free discontinuity problem’. *Numer. Funct. Anal. Optim.* **15**(3-4), 201–224.

11. Black, M. J., G. Sapiro, D. H. Marimont, and D. Heeger: 1998, 'Robust anisotropic diffusion'. *IEEE Trans. Image Processing* **7**(3), 421–432.
12. Blake, A. and A. Zisserman: 1987, *Visual reconstruction*. Cambridge, MA: MIT Press.
13. Blomgren, P. and T. F. Chan: 1998, 'Color TV: total variation methods for restoration of vector-valued images'. *IEEE Trans. Image Processing* **7**(3), 304–309.
14. Blomgren, P. V.: 1998, 'Total variation methods for restoration of vector valued images'. Ph.D. thesis, University of California, Los-Angeles.
15. Bonnet, A.: 1996, 'On the regularity of edges in image segmentation'. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **13**(4), 485–528.
16. Bonnet, A. and G. David: 2000, 'Cracktip is a global Mumford-Shah minimizer'. Technical Report 27, Prépublications d'Orsay.
17. Bourdin, B.: 1999, 'Image segmentation with a finite element method'. *M2AN Math. Model. Numer. Anal.* **33**(2), 229–244.
18. Bourdin, B. and A. Chambolle: 2000, 'Implementation of an adaptive finite-element approximation of the Mumford-Shah functional'. *Numer. Math.* **85**(4), 609–646.
19. Braides, A.: 1998, *Approximation of free-discontinuity problems*, Lecture Notes in Math., 1694. Berlin: Springer-Verlag.
20. Chambolle, A.: 1995, 'Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations'. *SIAM J. Appl. Math.* **55**(3), 827–863.
21. Chambolle, A.: 1999, 'Finite-differences discretizations of the Mumford-Shah functional'. *M2AN Math. Model. Numer. Anal.* **33**(2), 261–288.
22. Chambolle, A. and P.-L. Lions: 1997, 'Image recovery via total variation minimization and related problems'. *Numer. Math.* **76**, 167–188.
23. Chan, T. and J. Shen: 1999, 'Variational restoration of non-flat image features: models and algorithms'. Technical Report 99-20, UCLA CAM.
24. Chan, T. F., B. Y. Sandberg, and L. A. Vese: 2000, 'Active Contours without Edges for Vector-Valued Images'. *J. Visual Communication Image Representation* **11**(2), 130–141.
25. Chan, T. F. and L. A. Vese: 1999, 'An Active Contour Model without Edges'. In: M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert (eds.): *Scale-Space Theories in Computer Vision, Second International Conference*, Vol. 1682 of *Lecture Notes in Comp. Sci.* pp. 141–151.
26. Chan, T. F. and L. A. Vese: 2000, 'Image segmentation using level sets and the piecewise-constant Mumford-Shah model'. Technical Report 00-14, UCLA CAM.
27. Charbonnier, P., L. Blanc-Féraud, G. Aubert, and M. Barlaud: 1997, 'Deterministic edge-preserving regularization in computed imaging'. *IEEE Trans. Image Processing* **6**(2), 298–311.
28. Congedo, G. and I. Tamanini: 1991, 'On the existence of solution to a problem in multidimensional segmentation.'. *Ann. Inst. Henri Poincaré Anal. Non Linéaire* **8**(2), 175–195.
29. Cumani, A.: 1991, 'Edge detection in multispectral images'. *CVGIP: Graphical Models and Image Processing* **53**(1), 40–51.
30. Dacorogna, B.: 1989, *Direct methods in the calculus of variations*. Berlin: Springer-Verlag.
31. Dal Maso, G.: 1993, *An introduction to Γ -convergence*. Boston, MA: Birkhäuser Boston Inc.

32. De Giorgi, E.: 1991, 'Free discontinuity problems in calculus of variations'. In: R. Dautray (ed.): *Frontiers in pure and applied mathematics*. Amsterdam: North-Holland, pp. 55–62.
33. De Giorgi, E. and L. Ambrosio: 1988, 'New functionals in the calculus of variations'. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)* **82**(2), 199–210 (1989).
34. De Giorgi, E., M. Carriero, and A. Leaci: 1989, 'Existence theorem for a minimum problem with free discontinuity set'. *Arch. Rational Mech. Anal.* **108**(3), 195–218.
35. De Giorgi, E. and T. Franzoni: 1975, 'Su un tipo di convergenza variazionale'. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)* **58**(6), 842–850.
36. Di Zenzo, S.: 1986, 'A note on the gradient of a multi-image'. *Computer Vision, Graphics, and Image Processing* **33**(1), 116–125.
37. Dibos, F. and E. Séré: 1997, 'An approximation result for the minimizers of the Mumford-Shah functional'. *Boll. Un. Mat. Ital. A (7)* **11**(1), 149–162.
38. El-Fallah, A. I. and G. E. Ford: 1994, 'The evolution of mean curvature in image filtering'. In: *Proceedings of IEEE International Conference on Image Processing*, Vol. 1. pp. 298–302.
39. Evans, L. C. and R. F. Gariepy: 1992, *Measure theory and fine properties of functions*. Boca Raton, FL: CRC Press.
40. Focardi, M.: 2001, 'On the Variational Approximation of Free-Discontinuity Problems in the Vectorial Case'. *Math. Models and Methods in Appl. Sci.* **11**(4), 663–684.
41. Guseĭnov, F. V.: 1987, 'On the Jensen inequality'. *Math. Notes* **41**(5-6), 452–458.
42. Jost, J. and X. Li-Jost: 1998, *Calculus of Variations*. Cambridge University Press.
43. Kandel, E. R., J. H. Schwartz, and T. M. Jessell (eds.): 1991, *Principles of Neural Science*. Elsevier, third edition.
44. Kimmel, R., R. Malladi, and N. Sochen: 2000, 'Images as Embedded Maps and Minimal Surfaces: Movies, Color, Texture, and Volumetric Medical Images'. *Int. J. Computer Vision* **39**(2), 111–129.
45. Kimmel, R. and N. Sochen: 1999, 'Geometric-Variational Approach for Color Image Enhancement and Segmentation'. In: M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert (eds.): *Scale-space theories in computer vision*, Vol. 1682 of *Lecture Notes in Comp. Sci.* pp. 294–305.
46. Kimmel, R. and N. Sochen: 2000, 'Orientation diffusion or how to comb a porcupine'. *J. Visual Communication Image Representation*. accepted.
47. Koepfler, G., C. Lopez, and J.-M. Morel: 1994, 'A multiscale algorithm for image segmentation by variational method'. *SIAM J. Numer. Anal.* **31**(1), 282–299.
48. Lee, T. S., D. Mumford, R. Romero, and V. A. F. Lamme: 1998, 'The role of the primary visual cortex in higher level vision'. *Vision Research* **38**, 2429–2454.
49. Lee, T. S., D. Mumford, and A. Yuille: 1992, 'Texture segmentation by minimizing vector-valued energy functionals: the coupled-membrane model'. In: G. Sandini (ed.): *Computer vision – ECCV '92*, Vol. 558 of *LNCS*. pp. 165–173.
50. Lenz, R.: 2000, 'Color edge detectors for conical color spaces'. In: *International Conference on Color in Graphics and Image Processing*. pp. 284–289.
51. Li, S. Z.: 1995, *Markov random field modeling in computer vision*. Springer Verlag.

52. March, R. and M. Dozio: 1997, 'A variational method for the recovery of smooth boundaries'. *Image and Vision Computing* **15**, 705–712.
53. Marr, D.: 1982, *Vision*. San Francisco: W.H. Freeman.
54. Morel, J.-M.: 1997, 'The Mumford-Shah conjecture in image processing'. *Astérisque* (241), Exp. No. 813, 4, 221–242. Séminaire Bourbaki, Vol. 1995/96.
55. Morel, J.-M. and S. Solimini: 1995, *Variational methods in image segmentation*. Boston, MA: Birkhäuser Boston Inc. With seven image processing experiments.
56. Mumford, D.: 1994, 'Bayesian rationale for the variational formulation'. In: B. M. ter Haar Romeny (ed.): *Geometry-driven diffusion in computer vision*. Kluwer Academic Publishers, pp. 135–146.
57. Mumford, D. and J. Shah: 1985, 'Boundary detection by minimizing functionals'. In: *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 22–26.
58. Mumford, D. and J. Shah: 1989, 'Optimal approximations by piecewise smooth functions and associated variational problems'. *Comm. Pure Appl. Math.* **42**(5), 577–685.
59. Negri, M.: 1999, 'The anisotropy introduced by the mesh in the finite element approximation of the Mumford-Shah functional'. *Numer. Funct. Anal. Optim.* **20**(9-10), 957–982.
60. Osher, S. and L. I. Rudin: 1990, 'Feature-oriented enhancement using shock filters'. *SIAM J. Numer. Anal.* **27**(4), 919–940.
61. Richardson, T. and S. Mitter: 1994, 'Approximation, computation, and distortion in the variational formulation'. In: B. M. ter Haar Romeny (ed.): *Geometry-driven diffusion in computer vision*. Kluwer Academic Publishers, pp. 169–190.
62. Rosati, M.: 2000, 'Asymptotic behavior of a Geman and McClure discrete model'. *Appl. Math. Optim.* **41**(1), 51–85.
63. Rudin, L. I., S. Osher, and E. Fatemi: 1992, 'Nonlinear total variation based noise removal algorithms'. *Physica D* **60**, 259–268.
64. Samson, C., L. Blanc-Féraud, G. Aubert, and J. Zerubia: 2000a, 'Classification d'images multibandes par modèles variationnels'. Technical Report 4010, INRIA.
65. Samson, C., L. Blanc-Féraud, G. Aubert, and J. Zerubia: 2000b, 'A variational model for image classification and restoration'. *IEEE Trans. Pattern Analysis Machine Intelligence* **22**(5), 460–472.
66. Sapiro, G.: 1997, 'Color Snakes'. *Comput. Vision and Image Understanding* **68**(2), 247–253.
67. Sapiro, G. and D. L. Ringach: 1996, 'Anisotropic diffusion of Multivalued images with application to color filtering'. *IEEE Trans. Image Processing* **5**(11), 1582–1586.
68. Schnörr, C.: 1998, 'A study of a convex variational diffusion approach for image segmentation and feature extraction'. *J. Math. Imaging Vision* **8**(3), 271–292.
69. Shah, J.: 1994, 'Piecewise smooth approximations of functions'. *Calc. Var. Partial Differential Equations* **2**(3), 315–328.
70. Shah, J.: 1996a, 'A common framework for curve evolution, segmentation and anisotropic diffusion'. In: *CVPR '96: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 136–142.
71. Shah, J.: 1996b, 'Curve evolution and segmentation functionals: application to color images'. In: *Proceedings of 3rd IEEE International Conference on Image Processing*, Vol. 1. pp. 461–464.

72. Sochen, N., R. Kimmel, and R. Malladi: 1998, 'A general framework for low level vision'. *IEEE Trans. Image Processing* **7**(3), 310–318.
73. Sochen, N. and Y. Y. Zeevi: 1998, 'Representation of colored images by manifolds embedded in higher dimensional non-Euclidean space'. In: *Proceedings 1998 International Conference on Image Processing*, Vol. 1. pp. 166–170.
74. Spitaleri, R. M., R. March, and D. Arena: 1999, 'Finite difference solution of Euler equations arising in variational image segmentation'. *Numer. Algorithms* **21**(1-4), 353–365.
75. Tang, B., G. Sapiro, and V. Caselles: 2000, 'Diffusion of General Data on Non-Flat Manifolds via Harmonic Maps Theory: The Direction Diffusion Case'. *Int. J. Computer Vision* **36**(2), 149–161.
76. Teboul, S., L. Blanc-Féraud, G. Aubert, and M. Barlaud: 1998, 'Variational approach for edge-preserving regularization using coupled PDE's'. *IEEE Trans. Image Processing* **7**(3), 387–397.
77. Tsai, A., A. Yezzi, and A. S. Willsky: 2001, 'Curve evolution implementation of the the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification'. *IEEE Trans. Image Processing* **10**(8), 1169–1186.
78. Weickert, J.: 1998, *Anisotropic diffusion in image processing*. Stuttgart: B. G. Teubner.
79. Whitaker, R. and G. Gerig: 1994, 'Vector-valued diffusion'. In: B. M. ter Haar Romeny (ed.): *Geometry-driven diffusion in computer vision*. Kluwer Academic Publishers, pp. 93–134.
80. Yezzi, A.: 1998, 'Modified curvature motion for image smoothing and enhancement'. *IEEE Trans. Image Processing* **7**(3), 345–352.
81. You, Y.-L., W. Xu, A. Tannenbaum, and M. Kaveh: 1996, 'Behavioral Analysis of anisotropic diffusion in image processing'. *IEEE Trans. Image Processing* **5**(11), 1539–1553.