Chapter 4
The Mumford-Shah Problem in Computer Vision

Daniel Cremers and Bastian Goldlücke
Computer Vision Group
Technical University of Munich

Thomas Pock
Institute for Computer Graphics and Vision
Graz University of Technology
Overview

1. The Mumford-Shah problem
2. The piecewise constant case
3. The piecewise smooth case
Overview

1. The Mumford-Shah problem

2. The piecewise constant case

3. The piecewise smooth case
One of the first ideas for energy-minimization based image restoration/segmentation is due to [Geman, Geman ’84] and [Blake, Zisserman ’87]

Described in the discrete MRF context, where images \( g = (g_{i,j})_{1 \leq i \leq M, 1 \leq j \leq N}, g_{i,j} \in \{0, ..., 255\} \) are discrete 2D signals

In contrast to the TV-based edge-preserving regularization, they consider an additional edge indicator variable \( \ell = (\ell_{i+\frac{1}{2},j}, \ell_{i,j+\frac{1}{2}})^{i,j} \), which can take only values 0 or 1
One of the first ideas for energy-minimization based image restoration/segmentation is due to [Geman, Geman ’84] and [Blake, Zisserman ’87].

Described in the discrete MRF context, where images \( g = (g_{i,j})_{1 \leq i \leq M, 1 \leq j \leq N}, g_{i,j} \in \{0, \ldots, 255\} \) are discrete 2D signals.

In contrast to the TV-based edge-preserving regularization, they consider an additional edge indicator variable \( \ell = (\ell_{i+\frac{1}{2},j}, \ell_{i,j+\frac{1}{2}})_{i,j} \), which can take only values 0 or 1.

- \( \ell_{i+\frac{1}{2},j} = 1 \) \( \Rightarrow \) there is an edge between \((i,j)\) and \((i+1,j)\)
- \( \ell_{i+\frac{1}{2},j} = 0 \) \( \Rightarrow \) there is no edge
The approach of Geman and Geman

[Geman, Geman ’84] consider non-convex energies of the form

$$\min_{u,\ell} \sum_{i,j} \left[ (1 - \ell_{i+\frac{1}{2},j})(u_{i+1,j} - u_{i,j})^2 + (1 - \ell_{i,j+\frac{1}{2}})(u_{i,j+1} - u_{i,j})^2 \right] +$$

$$\mu \sum_{i,j} \left( \ell_{i+\frac{1}{2},j} + \ell_{i,j+\frac{1}{2}} \right) + \lambda \sum_{i,j} (u_{i,j} - g_{i,j})^2$$

Minimization is carried out using simulated annealing.

Image taken from [Geman, Geman ’84]
It has been observed by Mumford and Shah [Mumford, Shah ’89] that

- The set \( \{ \ell = 1 \} \) can be interpreted as a 1D curve \( K \subset \Omega \)
It has been observed by Mumford and Shah [Mumford, Shah ’89] that
• The set \( \{ \ell = 1 \} \) can be interpreted as a 1D curve \( K \subset \Omega \)
• The regularization with respect to \( \ell \) is proportional to its “length”
The Mumford-Shah problem

It has been observed by Mumford and Shah [Mumford, Shah ’89] that

- The set \( \{ \ell = 1 \} \) can be interpreted as a 1D curve \( K \subset \Omega \)
- The regularization with respect to \( \ell \) is proportional to its “length”
- They proposed to consider in a continuous setting, the minimal problem

\[
\min_{u,K} \mu \int_\Omega (u - g)^2 \, dx + \int_{\Omega \setminus K} |\nabla u|^2 \, dx + \nu \text{length}(K)
\]
The Mumford-Shah problem

It has been observed by Mumford and Shah [Mumford, Shah ’89] that

- The set \( \{ \ell = 1 \} \) can be interpreted as a 1D curve \( K \subset \Omega \)
- The regularization with respect to \( \ell \) is proportional to its “length”
- They proposed to consider in a continuous setting, the minimal problem

\[
\min_{u, K} \mu \int_{\Omega} (u - g)^2 \, dx + \int_{\Omega \setminus K} |\nabla u|^2 \, dx + \nu \text{length}(K)
\]

\( g \quad u \quad K \)
In a discrete setting no questions about the geometry of the minimizers can be raised.
Discrete versus Continuous

- In a discrete setting no questions about the geometry of the minimizers can be raised
- Existence of minimizers is a simple compactness theorem in finite dimensions
Discrete versus Continuous

- In a discrete setting no questions about the geometry of the minimizers can be raised
- Existence of minimizers is a simple compactness theorem in finite dimensions
- In the Mumford-Shah model, images are functions instead of vectors
Discrete versus Continuous

- In a discrete setting no questions about the geometry of the minimizers can be raised
- Existence of minimizers is a simple compactness theorem in finite dimensions
- In the Mumford-Shah model, images are functions instead of vectors
- This allows to study the geometry of minimizers
Discrete versus Continuous

- In a discrete setting no questions about the geometry of the minimizers can be raised
- Existence of minimizers is a simple compactness theorem in finite dimensions
- In the Mumford-Shah model, images are functions instead of vectors
- This allows to study the geometry of minimizers
- What does the minimal curve $K$ look like? Is it smooth?
Discrete versus Continuous

- In a discrete setting no questions about the geometry of the minimizers can be raised
- Existence of minimizers is a simple compactness theorem in finite dimensions
- In the Mumford-Shah model, images are functions instead of vectors
- This allows to study the geometry of minimizers
- What does the minimal curve $K$ look like? Is it smooth?
- The study of the Mumford-Shah problem has triggered a lot of interesting research in the past 30 years [Morel, Solimini '95], [David '99]
• In a discrete setting no questions about the geometry of the minimizers can be raised
• Existence of minimizers is a simple compactness theorem in finite dimensions
• In the Mumford-Shah model, images are functions instead of vectors
• This allows to study the geometry of minimizers
• What does the minimal curve $K$ look like? Is it smooth?
• The study of the Mumford-Shah problem has triggered a lot of interesting research in the past 30 years [Morel, Solimini ’95], [David ’99]
• The Mumford-Shah problem is both difficult to analyze mathematically and to solve numerically, since it requires to solve a non-convex problem.
Mumford and Shah conjectured that if the minimal segmentation is made of a finite set of smooth curves

\[ K = \bigcup_{i=1}^{N} K_i, \quad K_i \in C^1 \]
The Mumford-Shah conjecture

• Mumford and Shah conjectured that if the minimal segmentation is made of a finite set of smooth curves

\[ K = \bigcup_{i=1}^{N} K_i , \quad K_i \in C^1 \]

• \( K_i \) has the following local behavior
• Mumford and Shah conjectured that if the minimal segmentation is made of a finite set of smooth curves

\[ K = \bigcup_{i=1}^{N} K_i, \quad K_i \in C^1 \]

• \( K_i \) has the following local behavior

- \( C^1 \) curve
- triple junction
- crack tip
The Mumford-Shah conjecture

- Mumford and Shah conjectured that if the minimal segmentation is made of a finite set of smooth curves

\[ K = \bigcup_{i=1}^{N} K_i, \quad K_i \in C^1 \]

- \( K_i \) has the following local behavior

- Still unknown if completely true ...
Truncated quadratic regularization

- Consider the Mumford-Shah energy (without data fidelity)
  \[
  \int_{\Omega \setminus K} |\nabla u|^2 dx + \nu \text{length}(K)
  \]

- Is equivalent (in a spatially discrete setting) to truncated quadratic regularization where the truncation value corresponds to the parameter \(\nu\), [Chambolle ’95]
  \[
  \int_{\Omega} \phi_{\nu}(\nabla u) dx ,
  \]

  where
  \[
  \phi_{\nu}(\nabla u) = \min\{\nu, |\nabla u|^2\}
  \]
Continuation methods

Early attempts are mostly based on continuation methods

- Simulated annealing [Geman, Geman ’84]
  - Start with a high temperature $T$
  - Make random changes in the labels
  - Accept or reject changes with some probability depending on $T$
  - Gradually decrease $T$

- Graduated non-convexity (GNC) procedure [Blake, Zisserman ’87]

- Both methods are easy to implement and work reasonable well
The edge set can be substituted by using a family of non-decreasing functions $f : [0, +\infty) \rightarrow f : [0, +\infty)$ such that [Gobbino ’98], [Chambolle ’99]

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

Possible choices are $f(t) = \text{atan}(t)$ or $f(t) = \log(1 + t)$

- Works quite well in practice
Edge set $K$ is represented as the zero level set of a level set function $\phi$, [Osher, Sethian, ’88]

$$K = \{x \in \Omega : \phi(x) = 0\}$$

- Alternating minimization between $u$ and $K$ [Tsai, Yezzi, Willsky ’01], [Vese, Chan ’02], ...
- Works well for a certain class of applications
- Curve evolution step can also be solved globally using graph cuts, e.g. [Schoenemann, Cremers, ICCV ’07] [Grady, Alvino PAMI’09]
Application: Joint segmentation and registration
Phase field approximation

Phase fields approximation of [Ambrosio, Tortorelli, ’90]. The idea is to represent the edge set $K$ as a smooth edge indicator function $z : \Omega \rightarrow \mathbb{R}$ and designed the so-called phase-field energy

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

- The remarkable property: $L_{z, \varepsilon} \gamma$-converges to the length of $K$ as $\varepsilon \rightarrow 0^+$.
- Alternating minimization between $z$ and $u$
- Requires to solve linear systems only
- Works well in practice, but $\varepsilon$ has to be in the order of the grid size
The full Ambrosio-Tortorelli approximation is given by

$$\int_{\Omega} z^2 |\nabla u|^2 \, dx + \mu \left( \int_{\Omega} \varepsilon |\nabla z|^2 \, dx + \int_{\Omega} \frac{(1 - z)^2}{4\varepsilon} \, dx \right) + \lambda \int_{\Omega} (u - g)^2 \, dx$$
Consider again the Mumford-Shah functional

$$\min_{u,K} \int_{\Omega \setminus K} |\nabla u|^2 dx + \nu \text{length}(K) + \mu \int_{\Omega} (u - g)^2 dx$$
A limiting case

Consider again the Mumford-Shah functional

$$\min_{u,K} \int_{\Omega \setminus K} |\nabla u|^2 \, dx + \nu \text{length}(K) + \mu \int_{\Omega} (u - g)^2 \, dx$$

Now, restricting the function $u$ to piecewise constant functions, i.e. $u = a_l$, on each subset $E_l \in \Omega$ and multiplying the energy by $\mu^{-1}$, setting $\lambda = \nu / \mu$ yields the piecewise constant Mumford-Shah functional [Mumford, Shah '89]

$$\min_{a_l, K} \sum_l \int_{E_l} (a_l - g)^2 \, dx + \lambda \text{length}(K)$$
Consider again the Mumford-Shah functional

$$\min_{u,K} \int_{\Omega \setminus K} |\nabla u|^2 \, dx + \nu \text{length}(K) + \mu \int_{\Omega} (u - g)^2 \, dx$$

Now, restricting the function $u$ to piecewise constant functions, i.e. $u = a_l$, on each subset $E_l \in \Omega$ and multiplying the energy by $\mu^{-1}$, setting $\lambda = \nu / \mu$ yields the piecewise constant Mumford-Shah functional [Mumford, Shah '89]

$$\min_{a_l,K} \sum_l \int_{E_l} (a_l - g)^2 \, dx + \lambda \text{length}(K)$$

It is easy to see that the optimal $a_l$’s are given by the averages

$$a_l = \frac{\int_{E_l} g \, dx}{\int_{E_l} dx}$$
Consider again the Mumford-Shah functional

$$\min_{u,K} \int_{\Omega \setminus K} |\nabla u|^2 \, dx + \nu \text{length}(K) + \mu \int_{\Omega} (u - g)^2 \, dx$$

Now, restricting the function $u$ to piecewise constant functions, i.e. $u = a_l$, on each subset $E_l \in \Omega$ and multiplying the energy by $\mu^{-1}$, setting $\lambda = \nu/\mu$ yields the piecewise constant Mumford-Shah functional [Mumford, Shah '89]

$$\min_{a_l,K} \sum_l \int_{E_l} (a_l - g)^2 \, dx + \lambda \text{length}(K)$$

It is easy to see that the optimal $a_l$’s are given by the averages

$$a_l = \frac{\int_{E_l} g \, dx}{\int_{E_l} \, dx}$$

The piecewise constant case can also be viewed as the natural limit of the Mumford-Shah functional when $\mu \to 0$
1. The Mumford-Shah problem

2. The piecewise constant case

3. The piecewise smooth case
Interestingly, the Mumford-Shah conjecture proves to be true for the piecewise constant Mumford-Shah functional

- There exists an edge set $K$ made up of a finite number of $C^1$ arcs $K_i$ which join in a finite number of singular points (triple junctions in the interior of $\Omega$ and T-junctions on the boundary of $\Omega$)
Interestingly, the Mumford-Shah conjecture proves to be true for the piecewise constant Mumford-Shah functional

- There exists an edge set $K$ made up of a finite number of $C^1$ arcs $K_l$ which join in a finite number of singular points (triple junctions in the interior of $\Omega$ and T-junctions on the boundary of $\Omega$)

- Denoting $f_i(x) = (g(x) - a_i)^2$ it can be written as

$$\min_{E_i} \frac{1}{2} \sum_{l=1}^{k} \text{Per}(E_l; \Omega) + \sum_{l=1}^{k} \int_{E_l} f_i(x) \, dx,$$

$$\text{s.t. } \bigcup_{l=1}^{k} E_l = \Omega, \ E_s \cap E_t = \emptyset \forall s \neq t,$$

- Partitions the domain $\Omega \subset \mathbb{R}^d$ into $k$ pairwise disjoint sets $E_l$

- This is exactly the Ising (Potts) model in a continuous setting [Ising '25], [Potts '52]
The two-label case

A convex formulation for $k = 2$ has been presented in [Chan, Esedoglu, Nikolova ’06] by rewriting it in terms of the variational model. The idea is to introduce a labeling function $\theta : \Omega \rightarrow \{0, 1\}$, where $\theta(x) = 0$ if $x \in E_1$ and $\theta(x) = 1$ if $x \in E_2$

$$\min_{\theta} \int_{\Omega} |D\theta| + \int_{\Omega} (1 - \theta(x))f_1(x) + \theta(x)f_2(x) \, dx$$

Note that the total variation of a binary function corresponds to the boundary length of $E_1$

- Relaxation to $\theta \in BV(\Omega; [0, 1])$ yields a convex problem
- Application of the thresholding theorem allows to compute the global minimizer of the binary problem
The piecewise constant case

The multi-label case

- The discret version (Potts model) is known to be NP-hard
- A global solution is no longer possible
- The basic idea is to introduce $k$ relaxed labeling functions
  \[ \theta = (\theta_1, ... \theta_k) \in BV(\Omega; [0, 1]^k) \]
- Consider the following generic representation of the multi-label case

\[
\min_{\theta} J(\theta) + \sum_{l=1}^{k} \int_{\Omega} \theta_l f_l dx, \quad \text{s.t.} \quad \theta_l(x) \geq 0, \quad \sum_{l=1}^{k} \theta_l(x) = 1, \quad \forall x \in \Omega
\]
Convex relaxation

Different choices have been proposed

- The most straightforward relaxation has been proposed in [Zach, Gallup, Frahm, Niethammer ’08]

\[ \mathcal{J}_1(\theta) = \frac{1}{2} \sum_{l=1}^{k} \int_{\Omega} |D\theta_l| \]
Different choices have been proposed

- The most straightforward relaxation has been proposed in [Zach, Gallup, Frahm, Niethammer ’08]

\[ J_1(\theta) = \frac{1}{2} \sum_{l=1}^{k} \int_{\Omega} |D\theta_l| \]

- A more general version using a vectorial total variation has been proposed in [Lellmann, Kappes, Yuan, Becker, Schnörr ’08]

\[ J_2(\theta) = \int_{\Omega} \sqrt{\|D\theta_1\|_A^2 + \ldots + \|D\theta_k\|_A^2}, \quad \|v\|_A = \sqrt{v^T A^T A v} \]
Different choices have been proposed

- The most straightforward relaxation has been proposed in [Zach, Gallup, Frahm, Niethammer ’08]

$$J_1(\theta) = \frac{1}{2} \sum_{l=1}^{k} \int_{\Omega} |D\theta_l|$$

- A more general version using a vectorial total variation has been proposed in [Lellmann, Kappes, Yuan, Becker, Schnörr ’08]

$$J_2(\theta) = \int_{\Omega} \sqrt{\|D\theta_1\|_{A}^2 + \ldots + \|D\theta_k\|_{A}^2}, \|v\|_{A} = \sqrt{v^T A^T A v}$$

- A tighter relaxation using a local envelope approach has been proposed in [Chambolle, Cremers, Pock ’08]

$$J_3(\theta) = \int_{\Omega} \psi(D\theta), \quad \psi(p) = \sup_{q} \left\{ \sum_{l=1}^{k} \langle p_l, q_m \rangle : |p_l - q_m| \leq 1, 1 \leq l < m \leq k \right\}$$
A comparison using the “triple-junction” problem
A comparison using the “triple-junction” problem
Comparison

A comparison using the “triple-junction” problem
A comparison using the “triple-junction” problem
The piecewise constant case

Examples

The “4-label” problem

Note that the minimizer is composed of two “triple-junctions”
The “4-label” problem

Note that the minimizer is composed of two “triangle-junctions”
White/gray matter segmentation of the brain with $k = 4$ labels
Examples

Piecewise constant Mumford-Shah segmentation with $k = 10$ labels
The piecewise constant case

Examples

Piecewise constant Mumford-Shah segmentation with $k = 10$ labels

Piecewise constant Mumford-Shah segmentation with $k = 16$ labels
The “triple-junction” problem in 3D

One slice

3D rendering
Examples

Disparity estimation using the Potts model, simply use a different data term: \( f_l(x) = |l_{left}(x) - l_{right}(x + \text{disp}_l)| \)

Tsukuba data set, 64 labels, 300 it, 7.7s on a Tesla GPU
Disparity estimation using the Potts model, simply use a different data term: \( f_l(x) = |l_{\text{left}}(x) - l_{\text{right}}(x + \text{disp}_l)| \)

Tsukuba data set, 64 labels, 300 it, 7.7s on a Tesla GPU

Teddy data set, 256 labels, 300 it, 175.6s on a Tesla GPU
Overview

1. The Mumford-Shah problem

2. The piecewise constant case

3. The piecewise smooth case
The Euler-Lagrange equations of the Mumford-Shah functional provide only a necessary condition for minimality of a pair \((u, S_u)\) minimizing

\[
E(u) = \int_{\Omega \setminus S_u} |\nabla u|^2 dx + \nu \mathcal{H}^1(S_u) + \mu \int_{\Omega} (f - u)^2 dx
\]

In [Alberti, Bouchitte, Dal Maso, ’03], the authors provide a sufficient condition for (some) minimizers of the Mumford-Shah functional.

- The basic idea is to consider the complete graph of \(u\) instead of the function \(u\) only
- Then the idea is to rewrite the Mumford-Shah functional by means of the flux of a suitable vector field \(\varphi\) through the interface \(\Gamma_u\) of the subgraph
- In particular if for a particular \(u\), the vector field \(\varphi\) is found to be divergence-free then one can prove minimality of \(u\)
- The obtained vector field is called a calibration
- It is not clear whether for each minimizer there exists such a calibration
The piecewise smooth case

The calibration method

- The characteristic function $1_u$ of the subgraph of a function $u \in SBV(\Omega)$ is defined as

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

- The normal $\nu_{\Gamma_u}$ of the interface $\Gamma_u$ is given by

$$\nu_{\Gamma_u} = \begin{cases} 
\frac{(\nabla u, -1)}{\sqrt{|\nabla u|^2 + 1}}, & \text{if } u \in \Omega \setminus S_u \\
(\nu_u, 0), & \text{if } u \in S_u
\end{cases}$$
A gray value image and the interface of its subgraph
The piecewise smooth case

The calibration method

- Suppose, the maximum flux of a vector field \( \varphi \) through the interface \( \Gamma_u \) provides a lower bound to the Mumford-Shah energy

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

- It turns out that equality holds, if \( \varphi \) fulfills the following convex constraints

\[
K = \left\{ \varphi^t(x, t) \geq \frac{\varphi^x(x, t)^2}{4} - \mu(t - f(x))^2, \quad \left| \int_{t_1}^{t_2} \varphi^x(x, s) ds \right| \leq \nu \right\}
\]
A sufficient condition

- The integral can be extended to $\Omega \times \mathbb{R}$

\[
E(u) = \sup_{\varphi \in K} \int_{\Omega \times \mathbb{R}} \varphi D1_u,
\]

- The key observation is now: If for a given $u$ the supremum is attained by a divergence-free vector field $\varphi_u \in K$, one has

\[
E(v) = \sup_{\varphi \in K} \int_{\Omega \times \mathbb{R}} \varphi D1_v \geq \int_{\Omega \times \mathbb{R}} \varphi_u D1_v = \int_{\Omega \times \mathbb{R}} \varphi_u D1_u = E(u),
\]

- For any $v$ which agrees with $u$ on the boundary of $\Omega$
- Hence $u$ is a minimizer of the Mumford-Shah functional
- If the vector field is divergence-free, it is called a “calibration”
- It remains unclear if a calibration exists for each minimizer ...
Convex relaxation

- Relaxation of the binary function $1_u : \Omega \rightarrow \{0, 1\}$ to functions $\nu : \Omega \rightarrow [0, 1]$, $\lim_{t \rightarrow -\infty} \nu(x, t) = 1$, $\lim_{t \rightarrow +\infty} \nu(x, t) = 0$
- Results in the convex saddle-point problem [Pock, Cremers, Bischof, Chambolle '09]

\[
\min_{\nu} \left\{ \mathcal{E}(\nu) = \sup_{\varphi \in K} \int_{\Omega \times \mathbb{R}} \varphi D\nu \right\}
\]

- Solved using the primal dual algorithm [Pock, Cremers, Bischof, Chambolle '09], [Chambolle, Pock '10]
- The Euler-Lagrange equations imply that the optimal $\varphi$ is divergence free
- If the minimal $\nu$ is binary, the calibration argument can be applied
- Due to the non-local constraint in $K$ need to project on a huge convex set which is (very) time-consuming
- Yields high-quality solutions in most practical problems
- Note: Also works, if the data term is non-convex, e.g. stereo
Examples

The crack tip problem, optimality shown in [Bonnet, David ’01]
Examples

The crack tip problem, optimality shown in [Bonnet, David ’01]

Phase-field approximation

Convex relaxation/thresholding
The piecewise smooth case

Examples

Image restoration/segmentation

Phase-field approximation

Convex relaxation
Stereo example

Venus data set with the disparity space discretized into 20 labels, takes a long time to compute, even on the GPU ...
Summary and open questions

- Introduced the “famous” Mumford-Shah problem for computer vision
- Standard approaches to solve the Mumford-Shah problem
- Convex relaxations for the piecewise constant Mumford-Shah / Potts model
- Convex relaxation for the full Mumford-Shah model
- Applications
Summary and open questions

- Introduced the “famous” Mumford-Shah problem for computer vision
- Standard approaches to solve the Mumford-Shah problem
- Convex relaxations for the piecewise constant Mumford-Shah / Potts model
- Convex relaxation for the full Mumford-Shah model
- Applications

- Is there a calibration for each minimizer?
- Development of a faster algorithm?
- Extension to vector valued functions? [Mora ’02]