

# A Convex Representation for the Vectorial Mumford-Shah Functional

Evgeny Strekalovskiy<sup>1,2</sup>

Antonin Chambolle<sup>2</sup>

Daniel Cremers<sup>1</sup>

<sup>1</sup> Technical University Munich, Germany

<sup>2</sup> Ecole Polytechnique, Paris, France

## Abstract

We propose the first tractable convex formulation of the vectorial Mumford-Shah functional which allows to compute high-quality solutions independent of the initialization. To this end, we generalize recently introduced convex formulations for scalar functionals to the vector-valued scenario in such a way that discontinuities in the different color channels preferably coincide. Furthermore, we propose an efficient solution which makes the overall optimization problem as tractable as in the scalar-valued case. Numerous experimental comparisons with the naive channel-wise approach, with the well-known Ambrosio-Tortorelli approximation, and with the classical total variation confirm the advantages of the proposed relaxation for contrast-preserving and edge-enhancing regularization.

## 1. Introduction

### 1.1. The Mumford-Shah Problem

Regularization is of central importance in image analysis and beyond as it provides a prior for a number of otherwise ill-posed inverse problems. The Mumford-Shah functional [19] is a prototypical form of all regularizers which aim at combining a smoothing of homogeneous regions with the enhancement of edges: Given  $\Omega \subset \mathbb{R}^n$  a bounded open set, the vectorial Mumford-Shah problem is given by

$$\min_{u,K} \left\{ \int_{\Omega} |u - f|^2 dx + \alpha \int_{\Omega \setminus K} |\nabla u|^2 dx + \lambda |K| \right\} \quad (1)$$

where  $f : \Omega \rightarrow \mathbb{R}^k$  is the input image,  $\alpha, \lambda > 0$  are weights, and  $u = (u_1, \dots, u_k) : \Omega \rightarrow \mathbb{R}^k$  is the unknown which is assumed to be smooth except on a possible jump set  $K$ . This set is the same for all components  $u_i$  and introduces a coupling of the channels. The norm of the gradient  $|\nabla u|$  which appears in (1) is the Frobenius (euclidean) norm  $|\nabla u|^2 = \sum_i |\nabla u_i|^2$ , and the norm in the term  $|u - f|$  is also euclidean.



**Figure 1.** We propose a convex relaxation and an efficient implementation for the vectorial Mumford-Shah functional. It allows to compute piecewise smooth approximations of color images independent of initialization. Compared to the Total Variation (**center**), the Mumford-Shah model (**right**) preserves image contrast.

The Mumford-Shah problem has been intensively studied in the applied math community [18]. In practice its applicability is substantially limited because of its non-convexity. As a consequence, researchers typically revert to the (also non-convex) phase-field approximation of Ambrosio and Tortorelli [3, 4]. Extensions of this approximation to color images have been proposed in [7]. Alternatively it is often replaced with the convex total variation (TV). However, the tendency of TV to lower the contrast at edges and oversmooth flat regions (the “staircasing” phenomenon) makes it a poor substitute to more elaborate functionals—see Fig. 1. In [15] a different approach for color channels coupling is proposed by means of Riemannian Geometry.

### 1.2. Related Work

In the recent past, several authors have overcome the issue of non-convexity by suggesting convex relaxations for respective functionals [1, 11, 8]. Specific examples include the convex relaxation for the two-region piecewise-constant Mumford-Shah model [10], for multilabel problems with convex regularizer [22, 21], for the multi-region piecewise constant Mumford-Shah [16, 9, 24], and—possibly most closely related to this work—a convex relaxation for the *scalar* piecewise smooth Mumford-Shah model [20]. The key idea in convexifying multilabel problems is typically to derive from (1) a variational problem involving the entire graph of the function  $u$  (or, rather, the characteristic function of the sub-graph). This approach allows to naturally

incorporate the jump set  $K$ . In this new setting, respective problems can be well approximated by convex energies, which are relatively easy to solve (except, of course, that the complexity is increased). Some of these approaches were clearly inspired from the Markov random field (MRF) community where researchers have introduced graph cut algorithms for minimizing discrete version of such energies [12, 14, 6, 13]. In the MRF community, the Mumford-Shah regularizer corresponds to a truncated quadratic penalizer.

While the above works are predominantly focused on functionals over *scalar-valued* functions (greyscale images), the present work is focused on the case of multicomponent signals, i.e. color or multi-spectral images. The major challenge in formulating convex relaxations for vectorial models is that the straight-forward channel-by-channel application of respective scalar approaches [20] invariably produces suboptimal results because the individual color channels cannot be treated independently: Indeed, the jump set  $K$  in (1) combines all color channels, as it denotes all points in the image plane where the signal is discontinuous, no matter in which color channel.

A convex relaxation for the *vectorial* case has been given in [17] using the method of calibrations. However, it utilizes the full three-dimensional label space for color images and is therefore by no means tractable, neither in memory, nor in computation time. Moreover, it appears infeasible to actually devise a numerical algorithm for solving the relaxation of [17] because it depends only on a part of the solution.

### 1.3. Contributions

In this paper, we propose a convex representation for the vectorial Mumford-Shah functional (1) which captures the coupling among the different color channels correctly. It combines several important advantages:

- It is the first tractable convex relaxation for the vectorial Mumford-Shah functional. We propose an efficient implementation of the method for which both memory and runtime scale linearly with the number of channels. In particular, channel coupling is achieved with the same runtime as for the channel-wise version.
- The proposed method indeed favors solutions where discontinuities in the individual color channels coincide. As a consequence, in contrast to the channel-wise approach it does not introduce artificial color edges in the solution.
- In comparison to the channel-by-channel solution, the commonly employed Ambrosio-Tortorelli approximation [3, 4] and TV, the method leads to improved and more natural results for discontinuity-preserving denoising of color images and various other applications.

## 2. Preliminaries

Let  $\Omega \subset \mathbb{R}^n$  be a bounded open set. For vectorial functions  $u \in L^1(\Omega; \mathbb{R}^k)$  the *total variation* of  $u$  is defined by

$$TV(u) = \int_{\Omega} |Du| := \sup \left\{ \int_{\Omega} \sum_{i=1}^k u_i \operatorname{div} \varphi_i \, dx : \right. \\ \left. \varphi \in C_c^{\infty}(\Omega; (\mathbb{R}^n)^k), |\varphi(x)| \leq 1 \, \forall x \in \Omega \right\}. \quad (2)$$

The space of *functions of bounded variation*, i.e. where  $TV(u) < \infty$ , is denoted by  $BV(\Omega; \mathbb{R}^k)$  [2]. If  $u \in BV(\Omega; \mathbb{R}^k)$ , we denote by  $J_u$  its “essential” jump set. This is where one cannot unambiguously define  $u(x)$ , but rather two values  $u^- = (u_1^-, \dots, u_k^-)$  and  $u^+ = (u_1^+, \dots, u_k^+)$  and a normal vector  $\nu_u(x)$ , indicating the direction of the jump and pointing towards the  $u^+$  side. Note that when two or more components  $u_i$  jump at a point, the jump direction  $\nu_u(x)$  is indeed the same for  $\mathcal{H}^{n-1}$ -a.e.  $x$ . One says that  $u$  is a *special function of bounded variation*,  $u \in SBV(\Omega; \mathbb{R}^k)$ , if the distributional gradient can be decomposed as

$$Du = \nabla u \, dx + (u^+(x) - u^-(x)) \otimes \nu_u(x) \mathcal{H}^{n-1} \llcorner J_u \quad (3)$$

into a continuous part and a jump part, which is the jump times the Hausdorff  $(n-1)$ -dimensional measure restricted to the jump set  $J_u$ . Here,  $(u^+ - u^-) \otimes \nu_u$  denotes the matrix

$$((u_i^+ - u_i^-) \nu_u^j)_{1 \leq i \leq k, 1 \leq j \leq n}. \quad (4)$$

The “weak” Mumford-Shah functional is well-defined for functions in  $SBV(\Omega; \mathbb{R}^k)$  and given by

$$MS^{\alpha, \lambda}(u) = \int_{\Omega} |u - f|^2 \, dx + \alpha \int_{\Omega \setminus J_u} |\nabla u|^2 \, dx + \lambda \mathcal{H}^{n-1}(J_u). \quad (5)$$

## 3. The Convex Representation

### 3.1. Scalar Case

Given a function  $u : \Omega \rightarrow \mathbb{R}$ , its corresponding graph function  $\mathbf{1}_u : \Omega \times \mathbb{R} \rightarrow \{0, 1\}$  is defined as

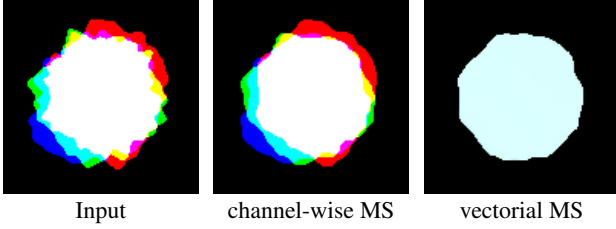
$$\mathbf{1}_u(x, t) = \begin{cases} 1 & \text{if } u(x) > t \\ 0 & \text{else.} \end{cases} \quad (6)$$

For the scalar case  $k = 1$ , the authors of [1] introduce the convex functional

$$\mathcal{F}(v) := \sup_{\sigma \in \mathcal{K}} \int_{\Omega \times \mathbb{R}} \sigma(x, t) \cdot Dv(x, t), \quad (7)$$

defined for  $v \in BV_{\text{loc}}(\Omega \times \mathbb{R})$ , where the constraint set is given by

$$\mathcal{K} := \left\{ \sigma \mid (\sigma^x, \sigma^t) \in C_c^{\infty}(\Omega \times \mathbb{R}; \mathbb{R}^n \times \mathbb{R}) \right. \\ \left. \sigma^t(x, t) \geq \frac{1}{4\alpha} |\sigma^x(x, t)|^2 - (t - f(x))^2, \right. \\ \left. \left| \int_t^{t'} \sigma^x(x, s) \, ds \right| \leq \lambda \quad \forall x \in \Omega, t < t' \right\}. \quad (8)$$



**Figure 2. Vectorial versus channel-wise Mumford-Shah.** The vectorial MS model penalizes jumps only once if two or more components jump at the same place. As a consequence, the correct representation (**right**) favors component edges to coincide leading to coherent segmentations. This is not the case for the channel-wise MS model (**center**) even for large weights  $\lambda$ .

It is then shown that for any  $u \in SBV(\Omega)$ ,

$$\mathcal{F}(\mathbf{1}_u) = MS^{\alpha, \lambda}(u). \quad (9)$$

The key advantage of this representation is that  $\mathcal{F}(v)$  is convex in  $v$ . This can be used to numerically minimize  $MS^{\alpha, \lambda}$  [20] with interesting practical applications.

### 3.2. Contribution: Vectorial Case

In the vectorial case  $k \geq 1$ , we consider  $k$  graph functions  $\mathbf{1}_u(x, t) = (\mathbf{1}_{u_1}, \dots, \mathbf{1}_{u_k})$  corresponding to the  $k$  channels  $u_i$ . We propose the following convex relaxation (in  $v = \mathbf{1}_u$ ) of the Mumford-Shah functional (5):

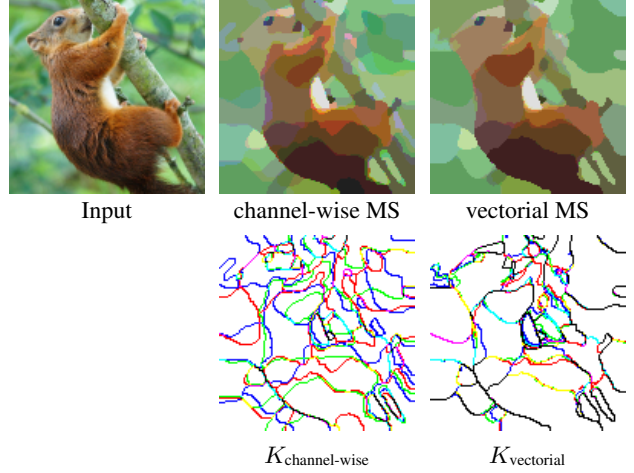
$$\mathcal{F}(v) := \sup_{\sigma \in \mathcal{K}} \sum_{i=1}^k \int_{\Omega \times \mathbb{R}} \sigma_i(x, t) \cdot Dv_i(x, t), \quad (10)$$

with the convex set

$$\begin{aligned} \mathcal{K} := \left\{ \sigma \mid (\sigma_i^x, \sigma_i^t) \in C_c^\infty(\Omega \times \mathbb{R}; \mathbb{R}^n \times \mathbb{R}), \right. \\ \sigma_i^t(x, t_i) \geq \frac{1}{4\alpha} |\sigma_i^x(x, t_i)|^2 - (t_i - f_i(x))^2, \\ \left. \sum_{j=1}^k \left| \int_{t_j}^{t'_j} \sigma_j^x(x, s) ds \right| \leq \lambda \right. \\ \left. \forall 1 \leq i \leq k, x \in \Omega, t_j < t'_j \right\}. \quad (11) \end{aligned}$$

The central part of the generalization is the second constraint of (11). Intuitively, the upper bound  $\lambda$  corresponds to the local penalization if each  $u_j$  jumps from  $t_j$  to  $t'_j$ . Let us explain how the set  $\mathcal{K}$  is derived. First, if  $u_i \in SBV(\Omega)$ , one can check (following the representation in [1]) that

$$\begin{aligned} \int_{\Omega \times \mathbb{R}} \sigma_i \cdot D\mathbf{1}_{u_i} \\ = \int_{\Omega} \sigma_i^x(x, u_i(x)) \cdot \nabla u_i(x) - \sigma_i^t(x, u_i(x)) dx \\ + \int_{J_{u_i}} \left( \int_{u_i^-(x)}^{u_i^+(x)} \sigma_i^x(x, s) ds \right) \cdot \nu_{u_i}(x) d\mathcal{H}^{n-1}(x). \quad (12) \end{aligned}$$



**Figure 3. Coupling of edge sets.** In contrast to the channel-by-channel Mumford-Shah (**center**), the vectorial Mumford-Shah model (**right**) favors discontinuities in respective channels to coincide. This is confirmed by the respective discontinuity sets  $K$  for each color component (**lower row**), where black lines represent edges in all three channels.

If  $\sigma_i$  satisfies the first constraint in (11), then standard convex duality shows that

$$\begin{aligned} \sigma_i^x(x, u_i(x)) \cdot \nabla u_i(x) - \sigma_i^t(x, u_i(x)) \\ \leq (u_i(x) - f_i(x))^2 + \alpha |\nabla u_i(x)|^2 \quad (13) \end{aligned}$$

a.e. in  $\Omega$ , with equality at  $x$  if and only if  $\sigma_i^t = |\sigma_i^x|^2 / (4\alpha) - (t - f_i)^2$  and  $t = u_i(x)$ . The right hand side of (13) summed up over all  $i$  are the first two integrands in (5). On the other hand, if  $\sigma$  satisfies the second constraint in (11), then one has a similar inequality for the jump part. Indeed, recalling that  $J_u = \bigcup_{i=1}^k J_{u_i}$ , one can check that

$$\begin{aligned} \sum_{i=1}^k \int_{J_{u_i}} \left( \int_{u_i^-(x)}^{u_i^+(x)} \sigma_i^x(x, s) ds \right) \cdot \nu_{u_i}(x) d\mathcal{H}^{n-1}(x) \\ \leq \int_{J_u} \sum_{i=1}^k \left| \int_{u_i^-(x)}^{u_i^+(x)} \sigma_i^x(x, s) ds \right| d\mathcal{H}^{n-1}(x) \quad (14) \\ \leq \lambda \mathcal{H}^{n-1}(J_u). \end{aligned}$$

This derivation shows that at least

$$\mathcal{F}(\mathbf{1}_u) \leq MS^{\alpha, \lambda}(u). \quad (15)$$

Beyond this inequality we will in fact show in Theorem 1 below that the proposed relaxation (10) indeed coincides with the original Mumford-Shah model (5). The proof will utilize a crucial efficient reformulation of the constraint set (11) which we introduce next.



**Figure 4. Piecewise smooth approximation without color artifacts.** A comparison on the Leonardo da Vinci’s “Dama con l’ermellino” shows that the channel-wise solution introduces independent discontinuities in the color channels, thereby producing colors not present in the original image (blue on the cheek and green in the hair for example). In contrast, the proposed convex relaxation of the vectorial Mumford-Shah model provides more natural color transitions.

### 3.3. Efficient Constraint Set Reformulation

In practice, the range of each channel  $u_i$  must be discretized into  $n_i \geq 1$  levels. In its original formulation (11), the constraint set  $K$  requires  $\mathcal{O}(n_1^2 \cdots n_k^2)$  constraints, which is not feasible in practice. However, it turns out that these can be equivalently reformulated using only  $\mathcal{O}(n_1^2 + \dots + n_k^2)$  constraints. Thus, the proposed relaxation (10) has the key advantage that its minimization is just as tractable as the simple channel-by-channel model. In Section 4 we will also propose an approximation which has even linear instead of quadratic complexity.

The idea of the constraint set decoupling is to introduce auxiliary variables  $m_i : \Omega \rightarrow \mathbb{R}$  for each  $1 \leq i \leq k$ .

**Proposition 1.** *The constraint set  $\mathcal{K}$  in (11) is equivalent to the following constraint set:*

$$\begin{aligned} \mathcal{K}' := & \left\{ (\sigma, m) \mid (\sigma_i^x, \sigma_i^t) \in C_c^\infty(\Omega \times \mathbb{R}; \mathbb{R}^n \times \mathbb{R}), \right. \\ & \sigma_i^t(x, t_i) \geq \frac{1}{4\alpha} |\sigma_i^x(x, t_i)|^2 - (t_i - f_i(x))^2, \\ & \left| \int_{t_i}^{t'_i} \sigma_i^x(x, s) ds \right| \leq m_i(x), \\ & \left. \sum_{j=1}^k m_j(x) \leq \lambda \quad \forall i, x \in \Omega, t_i < t'_i \right\}. \end{aligned} \quad (16)$$

*Proof.* Let  $(\sigma, m) \in \mathcal{K}'$ . Then obviously also  $\sigma \in \mathcal{K}$ . On the other hand, if  $\sigma \in \mathcal{K}$ , define  $m_i : \Omega \rightarrow \mathbb{R}$  by

$$m_i(x) := \sup_{t_i < t'_i} \left| \int_{t_i}^{t'_i} \sigma_i^x(x, s) ds \right|. \quad (17)$$

Then we have  $\sum_{j=1}^k m_j(x) \leq \lambda$  by the second inequality of (11), and together clearly  $(\sigma, m) \in \mathcal{K}'$ .  $\square$

**Theorem 1.** *Let  $u \in SBV(\Omega; \mathbb{R}^k)$ . Then*

$$\mathcal{F}(\mathbf{1}_u) = MS^{\alpha, \lambda}(u). \quad (18)$$

*Proof.* The inequality (15) has already been shown above, so it remains to show  $\mathcal{F}(\mathbf{1}_u) \geq MS^{\alpha, \lambda}(u)$ .

Consider  $k$  continuous functions  $m_i : \Omega \rightarrow \mathbb{R}$  for  $1 \leq i \leq k$  with  $\sum_{i=1}^k m_i(x) \leq \lambda$  for all  $x$  and  $\min_x m_i(x) > 0$  for all  $i$ . For each  $i$ , consider the non-uniform variant of the scalar Mumford-Shah functional:

$$\int_{\Omega} |u_i - f_i|^2 dx + \alpha \int_{\Omega \setminus J_{u_i}} |\nabla u_i|^2 dx + \int_{J_{u_i}} m_i(x) d\mathcal{H}^{n-1}(x), \quad (19)$$

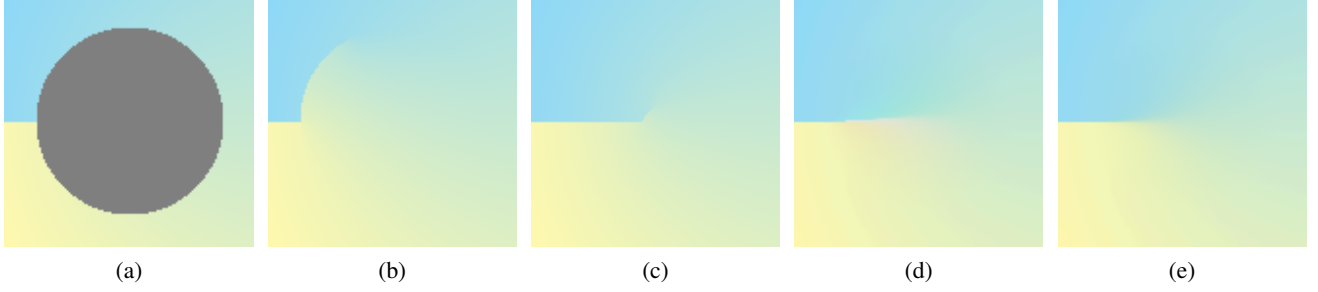
which is denoted by  $MS_i^{\alpha, m_i(\cdot)}(u_i)$ . Then, it can be quite easily shown that the vectorial MS functional (5) is

$$MS^{\alpha, \lambda}(u) = \sup_{m \in \mathcal{M}} \sum_{i=1}^k MS_i^{\alpha, m_i(\cdot)}(u_i) \quad (20)$$

where  $\mathcal{M}$  is the set of all  $m$ 's satisfying the above constraints. Moreover, for fixed  $m_i$ , an adaption of the proofs in [1, 20] will show that a representation similar to (9) holds, specifically

$$MS_i^{\alpha, m_i(\cdot)}(u_i) = \sup_{\sigma \in \mathcal{K}_i^{m_i(\cdot)}} \int_{\Omega \times \mathbb{R}} \sigma(x, t) \cdot D\mathbf{1}_{u_i}(x, t) \quad (21)$$

for  $u_i \in SBV(\Omega)$ , where  $\mathcal{K}_i^{m_i(\cdot)}$  is defined as in the scalar case (8) but with  $\lambda$  replaced by  $m_i(x)$ . Theorem 1 then follows from (20) and (21) because  $\sup_{\mathcal{K}'} \geq \sup_{m \in \mathcal{M}} \sup_{\sigma_i \in \mathcal{K}_i^{m_i(\cdot)}}$ , as  $m$  is more general in (16).  $\square$



**Figure 5. Inpainting test case: the cracktip problem.** (a) Grey inpainting domain and colors prescribed on the boundary. (b) Ambrosio-Tortorelli method can get stuck in bad local minima. (c) Ambrosio-Tortorelli method for a good initial guess. (d) Channel-wise MS introduces faulty red colors in the center. (e) Vectorial MS produces a good approximation to the cracktip solution.

### 3.4. Generalizations and Variants

One can generalize the Mumford-Shah model (5), replacing the data term  $|u - f|^2$  and the regularization term  $\alpha|\nabla u|^2$  with separable terms of the form  $\sum_{i=1}^k \varrho_i(x, u_i(x))$  and  $\sum_{i=1}^k \varphi_i(\nabla u_i)$  with convex  $\varphi_i$ , respectively. This amounts to replacing the first constraint in (16) with

$$\sigma_i^t(x, t) \geq \varphi_i^*(\sigma_i^x(x, t)) - \varrho_i(x, t) \quad (22)$$

for all  $i, x, t$ . Here,  $\varphi_i^*(q) := \sup_p p \cdot q - \varphi_i(p)$  is the Legendre-Fenchel conjugate of  $\varphi_i$ . Then, provided  $\varrho_i$  is continuous, and  $\varphi_i$  is convex and superlinear, i.e.  $\lim_{t \rightarrow \infty} \varphi_i(tp)/t = \infty$  for  $p \neq 0$ , Theorem 1 will hold.

For example, the data term  $\varrho_i(x, u_i) = |u_i - f_i|$  is suited to remove salt-and-pepper noise, while  $|u - f|^2$  is applicable for Gaussian noise. More generally, one is free to choose possibly *non-convex* data terms. With  $\varphi_i(\nabla u_i) = \alpha_i |\nabla u_i|^2$  one can choose different gradient weightings  $\alpha_i > 0$ .

### 4. Implementation

In order to minimize  $\mathcal{F}(1_u)$  and thus  $MS^{\alpha, \lambda}(u)$ , we use the convexity of  $\mathcal{F}$  and instead directly minimize over the graph function  $v = 1_u$ . The overall problem becomes

$$\inf_{v \in \mathcal{C}} \sup_{(\sigma, m) \in \mathcal{K}'} \sum_{i=1}^k \int_{\Omega \times \mathbb{R}} \sigma_i(x, t) \cdot Dv_i(x, t). \quad (23)$$

The constraint set for the solutions

$$\mathcal{C} := \left\{ v \in BV_{loc}(\Omega \times \mathbb{R}; [0, 1])^k \mid \begin{aligned} v_i(x, t) &= 1 \quad \forall t \leq a_i, \quad v_i(x, t) = 0 \quad \forall t > b_i, \\ v_i(x, \cdot) &\text{ non-increasing} \end{aligned} \right\} \quad (24)$$

is the convex hull of the valid graph functions, i.e. we relax the binary constraint  $v_i(x, t) \in \{0, 1\}$  allowing values in-between. The scalars  $a_i < b_i$  define the range of the channels  $u_i : \Omega \rightarrow \Gamma_i := [a_i, b_i]$ . Theorem 1 assures that

if we have a binary solution  $v = 1_u$  of (23), then  $u$  must indeed be a global optimum of the vectorial MS model (5).

Discretizing the domain  $\Omega$  and the channel range sets  $\Gamma_i$ , we solve the saddle-point problem (23) by the fast primal-dual algorithm of [20]. This is basically a gradient descent in  $v$  and gradient ascent in  $\sigma, m$ , with reprojections onto the constraint sets. For the projection onto  $\mathcal{C}$  in (24), observe that  $\sigma_i^t$  is allowed to be arbitrarily large in (16) and thus the supremum in (23) already ensures  $\partial_t v_i(x, t) \leq 0$ , i.e. that  $v_i(x, \cdot)$  is non-increasing. Other constraints are pointwise.

To project onto  $\mathcal{K}'$  in (16), observe that introducing auxiliary variables  $p_i : \Omega \times \Gamma_i^n \rightarrow \mathbb{R}^n$  with  $\partial_t p_i(x, t) = \sigma_i^x(x, t)$  for all  $x, t$ , the second constraint in (16) becomes

$$|p_i(x, t') - p_i(x, t)| \leq m_i(x). \quad (25)$$

Finally, we enforce both conditions on  $p$  using Lagrange multipliers and convex duality, adding the terms

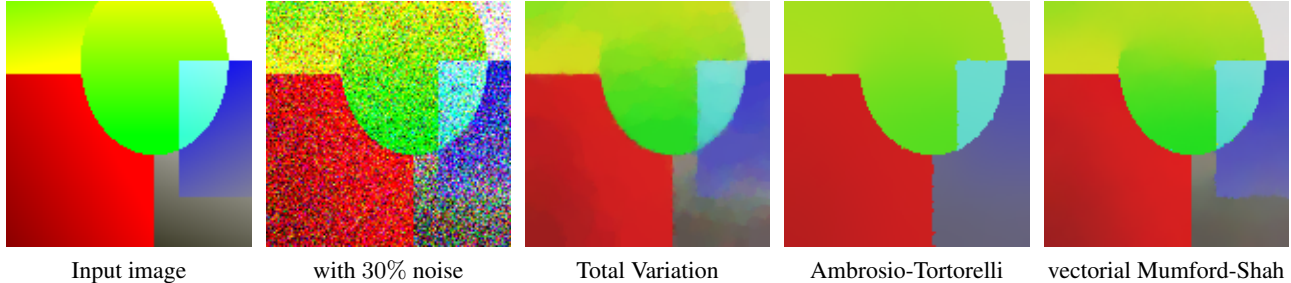
$$\begin{aligned} \inf_{\xi, \eta} \int_{\Omega} \sum_{i, t < t'} \xi_i(x, t, t') (p_i(t') - p_i(t)) + \eta_i(x, t, t') m_i(x) \\ + \inf_{\mu} \int_{\Omega} \sum_{i, t} \mu_i(x, t) (\partial_t p_i(x, t) - \sigma_i^x(x, t)) \end{aligned} \quad (26)$$

to the energy, with auxiliary variables  $\mu : \Omega \times \Gamma_i \rightarrow \mathbb{R}^n$  and  $(\xi, \eta) : \Omega \times \Gamma_i \times \Gamma_i \rightarrow \mathbb{R}^n \times \mathbb{R}$  such that pointwise  $|\xi| \leq \eta$ . The overall saddle-point problem (23) is then minimized over  $v, \xi, \eta, \mu$  and maximized over  $\sigma, m, p$ .

**Linear Complexity Approximation.** The quadratically many constraints (25) are satisfied if the stricter condition

$$|p_i(x, t)| \leq \frac{1}{2} m_i(x) \quad \forall t \in \Gamma_i \quad (27)$$

holds for all  $x \in \Omega$  and  $1 \leq i \leq k$ . These constraints can be implemented as in the first line of (26). Using (27) instead of (25) yields a less tight relaxation of the Mumford-Shah functional. However, the time and memory complexities are reduced substantially from quadratic to just *linear*. In the special case  $k = 1$  and  $\alpha = \infty$ , i.e. scalar partitioning, the approximation (27) corresponds to the relaxation [24].



**Figure 6. Denoising test case.** Total Variation leads to a loss of contrast and staircasing (piecewise constant regions) at the lower right corner. The Ambrosio-Tortorelli method optimizes only locally, missing the blue region. In contrast, the proposed convex relaxation of the vectorial Mumford-Shah model provides the best reconstruction.

## 5. Experimental Results

In the following, we will provide experimental comparisons of the proposed vectorial Mumford-Shah relaxation with several alternative algorithms on a variety of inverse problems. The channels  $u_i$  are obtained from the computed graph functions  $v_i$  in (6) by taking the 0.5-isolevel. When comparing channel-wise and vectorial MS, we set  $\lambda_{\text{channel-wise}} = \lambda_{\text{vectorial}}/3$ , so that both functionals are two different convex relaxations of the same energy (5), since channel-wise MS counts common boundaries up to three times.

On NVIDIA GTX 480, a typical runtime for  $128 \times 128$  color images with 32 levels for each color channel is about 20 seconds. For comparison, the channel-wise version runs in 19, the proposed relaxation with the simplified constraint set (27) in 3.5, Ambrosio-Tortorelli approximation in 1, and TV in 0.02 seconds. Thus, the proposed method accounts for the channel coupling while remaining as efficient as the simple channel-by-channel version.

### 5.1. Piecewise Smooth Approximations

Fig. 2 shows a synthetic  $128 \times 128$  image with three different blobs for each color channel. In the piecewise smooth approximation, the vectorial MS model clearly favors solutions where the edge sets coincide. This is not the case for the channel-wise variant, which processes one color at a time and is thus “color blind” w.r.t. color as a whole. This is further confirmed in Fig. 3 on a real world image by directly visualizing the edge sets of the different colors (coloring them accordingly). We used a  $8 \times 8 \times 8$  color discretization.

In Fig. 4 we compute piecewise smooth approximations for the  $256 \times 256$  image “Dama con l’ermellino” by Leonardo da Vinci using vectorial MS and channel-wise MS. The parameters are  $\alpha = 100$  and  $\lambda = 0.1$ , with a  $32 \times 32 \times 32$  discretization. The large parameter  $\alpha$  makes the approximation nearly piecewise constant, showing significant color artifacts of the channel-wise MS model, as opposed to the vectorial one.

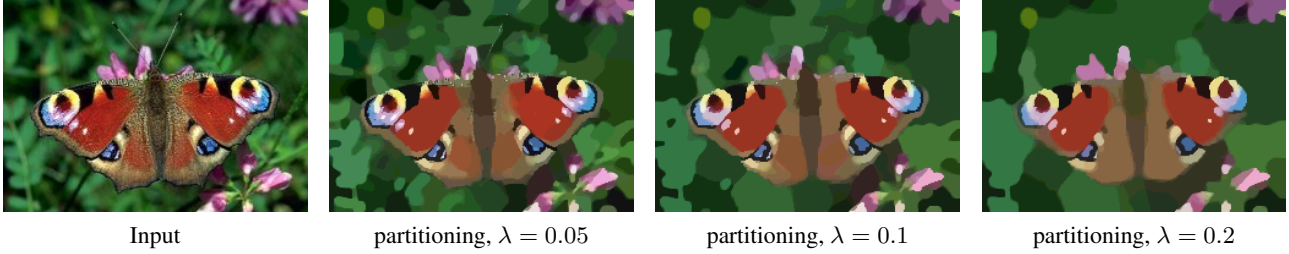
### 5.2. Denoising

To compare various regularizers, we devised a synthetic denoising test case in Fig. 6, degrading the  $128 \times 128$  test image by adding severe 30% noise. The result using simple total variation regularization produces loss of contrast and staircasing effects are visible in the grey lower right part. The result of the Ambrosio-Tortorelli method [4] is also shown. It is a non-convex approximation of the Mumford-Shah energy and thus merely allows to compute a local minimum, missing the blue region. Finally, the proposed vectorial MS relaxation provides more natural results than the other approaches ( $32 \times 32 \times 32$ ). Fig. 1 further demonstrates the loss of contrast of TV on a real world image.

### 5.3. The Cracktip Problem

Among the fascinating aspects of the Mumford-Shah functional is that it allows to model open boundaries. In the scalar case, for inpainting a circular domain with a Mumford-Shah regularizer and specific values prescribed on the boundary, a well-known analytical solution is the so-called cracktip function [5], in polar coordinates  $f(r, \varphi) = c\sqrt{r} \sin \frac{\varphi}{2}$  for some  $c > 0$ . In the vectorial case, applying different linear transformations to  $f$  for each color channel, one can easily see that the “color cracktips” are also solutions of the MS problem. It is thus interesting to see, how well this solution is recovered with different algorithms.

Fig. 5 shows the corresponding  $128 \times 128$  inpainting experiment using a  $64 \times 64 \times 64$  color discretization. We set the dataterm to  $\varrho = 0$  inside the circle and to  $M(u - f)^2$  outside with a big  $M > 0$ . As expected, the non-convex Ambrosio-Tortorelli approximation greatly relies on the initial solution to produce acceptable results: With a bad initialization it generates strong artifacts (b). Yet, even with a good initial guess, the same artifact emerges at a smaller scale (c). The channel-wise MS model treats the color channels independently, and thus introduces faulty red colors in the center (d). In contrast, the proposed convex relaxation of the vectorial MS model provides a good approximation of the color cracktip (e).



**Figure 7. Joint segmentation and color selection.** In contrast to previous relaxations where the set of allowed color models must be specified beforehand, the proposed approach automatically selects the appropriate color models during the segmentation process based on the scale parameter  $\lambda$ . Color discretization is  $16 \times 16 \times 16$ .

### 5.4. Unsupervised Image Partitioning

In the limiting case  $\alpha \rightarrow \infty$  in (5), the smoothness constraint is enforced more and more leading to the well-known *piecewise constant* approximation, also known as the *minimal partition problem*. The first constraint in (16) reduces to just  $\sigma_i^t(x, t) \geq -(t - f_i(x))^2$ . Only one parameter  $\lambda$  remains in the model (5), which controls the overall length of the interfaces between the regions where  $u$  is constant.

The standard approach to compute such a multi-region partitioning is by alternatingly determining a small number of color models (with fixed regions) and optimizing for the regions (with fixed color models). Of course, for such iterative approaches results are invariably suboptimal and will depend on the choice of the initial color models.

In contrast, the proposed convex relaxation allows to *jointly* optimize over the regions of constancy and the corresponding colors in these regions in a single convex optimization problem. In particular, the algorithm determines the appropriate number of color models depending on the input image and the scale parameter  $\lambda$ . Fig. 7 shows the partitionings computed for various values of  $\lambda$ .

### 5.5. Joint Disparity and Segmentation

Our final experiment is a more advanced application of the vectorial Mumford-Shah model (5) to the stereo image analysis. Given a stereo image pair, the task is to jointly compute a disparity map and a color segmentation, the central idea being that discontinuities in disparity and color tend to coincide. While this problem has been addressed in the non-convex Ambrosio-Tortorelli framework [23], we can apply the proposed convex relaxation of the vectorial Mumford-Shah functional to compute solutions independent of initialization.

The proposed problem simply corresponds to the case of  $k = 4$  channels corresponding to the three color and one depth channel:  $u = (u^{\text{color}}, u^{\text{depth}})$ . For the data term and gradient penalization we use the generalized variants of

Section 3.4. We use the data term from [23]:

$$\begin{aligned} \varrho_i^{\text{color}}(x, t) &= (1 - \gamma) (t - I_{\text{left}}^i(x))^2, \\ \varrho_i^{\text{depth}}(x, t) &= \gamma \sum_{j=1}^D \frac{(t - d_j(x))^2}{1 + (t - d_j(x))^2} \end{aligned} \quad (28)$$

with  $\gamma = 0.05$  and  $D = 4$  depth hypotheses  $d_j : \Omega \rightarrow \mathbb{R}$  calculated as in [23, Section 5.1]. Furthermore, we set  $\varphi_i^{\text{color}}(\nabla u_i) = 2|\nabla u_i|^2$  and  $\varphi_i^{\text{depth}}(\nabla u) = 100|\nabla u|^2$ .

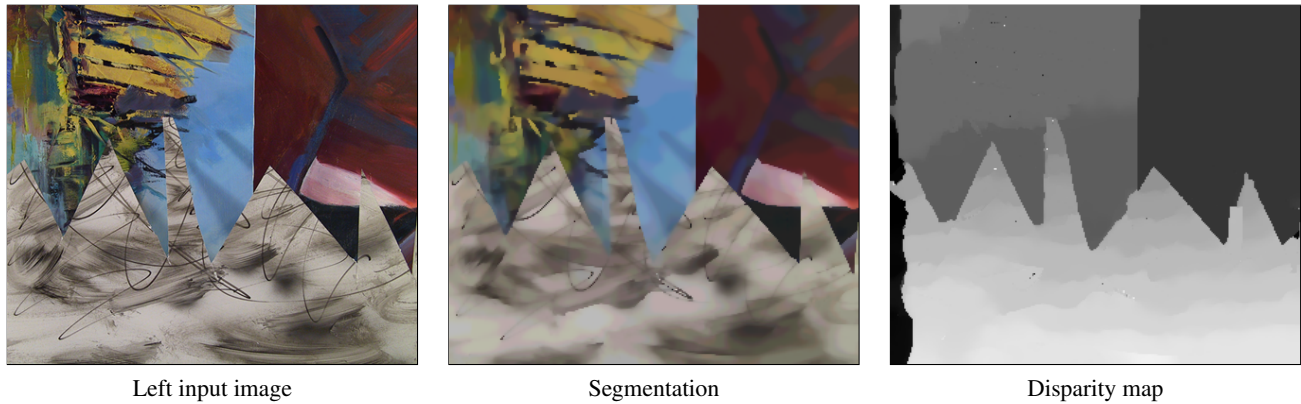
Fig. 8 shows the jointly computed segmentation and disparity for the  $289 \times 253$  “sawtooth” test image of the Middlebury stereo dataset. We set  $\lambda = 0.01$ , using a  $16 \times 16 \times 16 \times 20$  discretization. The clear correlation between the color and disparity edges, especially at the right (non-occluding) boundaries of the sawtooths confirms the advantage of the proposed model.

## 6. Conclusion

We proposed a convex relaxation for the vectorial Mumford-Shah problem. In contrast to a naive sequential processing of each color channel, this approach allows to correctly handle the coupling of all color channels with a single discontinuity set. As a consequence, it assures that color discontinuities tend to coincide and thus avoids color artifacts. Furthermore, the relaxation is computationally tractable. We proposed an efficient algorithmic implementation which allows to capture the color coupling at no additional runtime. Experimental comparisons with the channel-wise application of the scalar Mumford-Shah model, with total variation and with the non-convex Ambrosio-Tortorelli approach confirm the superiority of the proposed Mumford-Shah relaxation for contrast-preserving and edge-enhancing regularization in vectorial inverse problems.

## References

- [1] G. Alberti, G. Bouchitté, and G. Dal Maso. The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calc. Var. Partial Differential Equations*, 16(3):299–333, 2003. 1, 2, 3, 4



**Figure 8. Joint segmentation and stereo depth reconstruction.** By applying the convex proposed relaxation of the vectorial Mumford-Shah model to the joint estimation of color segmentation and disparity, we can impose that discontinuities of the color values coincide with depth discontinuities. In this setting, the label space is four dimensional with three color channels and one depth channel.

- [2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, USA, 2000. 2
- [3] L. Ambrosio and V. M. Tortorelli. Variational problems in SBV and image segmentation. *Acta Applicandae Mathematicae*, 17(1):1–40, 1989. 1, 2
- [4] L. Ambrosio and V. M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via  $\Gamma$ -convergence. *Comm. Pure Appl. Math.*, 43:999–1036, 1990. 1, 2, 6
- [5] A. Bonnet and G. David. Cracktip is a global Mumford-Shah minimizer. *Astérisque*, (274):vi+259, 2001. 6
- [6] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 23(11):1222–1239, 2001. 2
- [7] A. Brook, R. Kimmel, and N. A. Sochen. Variational restoration and edge detection for color images. *Journal of Mathematical Imaging and Vision*, 18(3):247–268, 2003. 1
- [8] A. Chambolle. Convex representation for lower semicontinuous envelopes of functionals in  $L^1$ . *J. Convex Anal.*, 8(1):149–170, 2001. 1
- [9] A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Technical report TR-2008-05, Dept. of Computer Science, University of Bonn, Bonn, Germany, November 2008. 1
- [10] T. Chan, S. Esedoğlu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006. 1
- [11] M. Giaquinta, G. Modica, and J. Souček. *Cartesian currents in the calculus of variations I, II.*, volume 37-38 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3*. Springer-Verlag, Berlin, 1998. 1
- [12] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *J. Roy. Statist. Soc., Ser. B.*, 51(2):271–279, 1989. 2
- [13] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, 2003. 2
- [14] H. Ishikawa and D. Geiger. Segmentation by grouping junctions. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 125–131, 1998. 2
- [15] R. Kimmel, R. Malladi, and N. A. Sochen. Image processing via the Beltrami operator. In *ACCV*, volume 1351, pages 574–581. Springer, 1998. 1
- [16] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. Technical report, IPA, HCI, Univ. Heidelberg, October 2008. 1
- [17] M. G. Mora. The calibration method for free-discontinuity problems on vector-valued maps. *J. Convex Anal.*, 9(1):1–29, 2002. 2
- [18] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhäuser, Boston, 1995. 1
- [19] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42(5):577–685, 1989. 1
- [20] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In *ICCV*, 2009. 1, 2, 3, 4, 5
- [21] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM Journal on Imaging Sciences*, 3(4):1122–1145, 2010. 1
- [22] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *ECCV*, Marseille, France, October 2008. 1
- [23] T. Pock, C. Zach, and H. Bischof. Mumford-Shah meets stereo: Integration of weak depth hypotheses. In *CVPR*, 2007. 7
- [24] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Workshop on Vision, Modeling and Visualization*, October 2008. 1, 5