

A Convex Relaxation Approach for Computing Minimal Partitions*

Thomas Pock

Graz University of Technology, Austria

pock@icg.tugraz.at

Daniel Cremers

Bonn University, Germany

dcremers@cs.uni-bonn.de

Antonin Chambolle

Ecole Polytechnique, Paris, France

antonin.chambolle@polytechnique.fr

Horst Bischof

Graz University of Technology, Austria

bischof@icg.tugraz.at

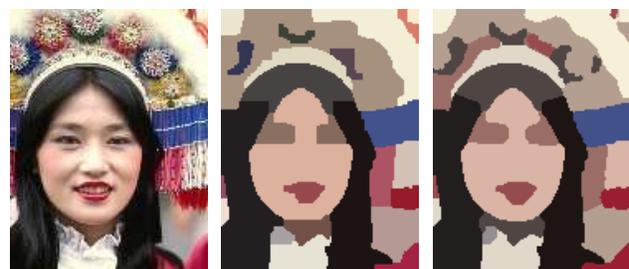
Abstract

In this work we propose a convex relaxation approach for computing minimal partitions. Our approach is based on rewriting the minimal partition problem (also known as Potts model) in terms of a primal dual Total Variation functional. We show that the Potts prior can be incorporated by means of convex constraints on the dual variables. For minimization we propose an efficient primal dual projected gradient algorithm which also allows a fast implementation on parallel hardware. Although our approach does not guarantee to find global minimizers of the Potts model we can give a tight bound on the energy between the computed solution and the true minimizer. Furthermore we show that our relaxation approach dominates recently proposed relaxations. As a consequence, our approach allows to compute solutions closer to the true minimizer. For many practical problems we even find the global minimizer. We demonstrate the excellent performance of our approach on several multi-label image segmentation and stereo problems.

1. Introduction

Many Computer Vision problems can be formulated as labeling problems, where each pixel in the image is assigned one of several labels. The major challenge is to find an optimal label configuration for all image pixels among all possible configurations. Since most Computer Vision tasks are highly ill-posed [15], optimal solutions are typically determined as the lowest energy configurations of some objective function that appropriately describes the respective Computer Vision problem. The energy of a configuration

*This work was supported by the Hausdorff Center for Mathematics, the Austrian Research Promotion Agency within the VM-GPU project (no. 813396) and the Austrian Science Fund (FWF) under the doctoral program *Confluence of Vision and Graphics* W1209. We also greatly acknowledge Nvidia for providing us a Tesla C1060 GPU.



(a) Input image (b) Discrete MRF (c) Proposed method

Figure 1. Piecewise constant Mumford-Shah segmentation of a natural color image using the proposed approach. (a) shows the input image, (b) is the result of a discrete MRF formulation using alpha expansion (image is courtesy of D. Mumford [23]). (c) shows the result with the proposed spatially continuous formulation. One can see that in contrast to the discrete MRF approach, the proposed method produces spatially consistent boundaries.

serves as a measure of its quality, better solutions correspond to lower energies. It is well known that the process of energy minimization is equivalent to a Bayesian maximum a posteriori estimation, where the prior probability gives rise to corresponding regularity terms in the energy. Labels can either be defined on a discrete grid or in a spatially continuous domain. In the discrete setting, respective optimization problems can be addressed on a graph with nodes and edges in the framework of *Markov Random Field* (MRF) theory [19]. If the set of possible labels is binary, and the pairwise terms are submodular, combinatorial algorithms such as graph cuts can be used to compute global minimizers [14, 21]. Applications of binary labeling problems include two-label image segmentation, shape denoising and 3D reconstruction.

1.1. Multi-Label Problems in Computer Vision

A substantially larger class of Computer Vision problems, including image restoration, inpainting, multi-region

image segmentation, motion and stereo, can be cast as multi-label optimization problems. Unfortunately, most multi-label problems cannot be minimized globally. They can only be solved approximately either by transforming the problem to a sequence of binary labeling problems [5, 28], via various *linear programming relaxations* [3, 29] or via roof duality relaxation [16] which has recently attracted renewed interest in the MRF community (cf. [27]). A notable exception is the work of Ishikawa [17], who showed that *exact* solutions for certain multi-label problems can be computed in polynomial time as binary cuts of higher-dimensional graphs provided that the pairwise interaction terms are convex with respect to a linearly ordered label set. Unfortunately, many interesting Computer Vision problems are not covered by the class considered by Ishikawa. First, in order to preserve discontinuities in the solution, researchers favor energies with *non-convex* regularizers. Second, many multi-labeling problems in Computer Vision do not provide an ordering on the space of labels (e.g. segmentation). Nevertheless, one can impose a regularity which favors labels of neighboring points to be identical. The resulting energy is known as the *Potts model* [26]. How to efficiently compute solutions for the spatially continuous Potts model is the focus of this paper.

Much work has been done to tackle multi-label problems in the discrete setting but only little work has been done on optimal solutions for labeling problems defined in a spatially continuous domain. The major difference between the discrete and the continuous setting is, that in the continuous setting, the label configuration is replaced by a labeling function. Hence, the discrete energy is now given by an energy functional. Here, the *calculus of variations* provides the suitable mathematical theory. Its fundamental theorem, the Euler-Lagrange differential equation, provides a *necessary* condition to describe the functional at stationary points (e.g. minimizers). The clear advantage of the spatially continuous formulation is, that it does not suffer from grid bias, also known as metrication error [20]. Notable examples in the continuous setting are the level set methods which allow to compute *locally optimal* solutions (cf. [13]), convex relaxation and thresholding techniques which allow to compute optimal binary labelings [11], optimal continuous multi-labelings in the case of convex regularity terms [25] and calibration methods [1, 8].

1.2. The Potts Model

The Potts model [26] has its origin in statistical mechanics to model phenomena of solid state physics. It is the generalization of the two-state model of Ising [18] to multiple states (labels). A spatially continuous formulation of

the Potts model is the partitioning problem

$$\min_{E_l} \left\{ \frac{1}{2} \sum_{l=0}^k \text{Per}(E_l; \Omega) + \sum_{l=0}^k \int_{E_l} f_l(x) dx \right\}, \quad (1)$$

such that $\bigcup_{l=0}^k E_l = \Omega, \quad E_s \cap E_t = \emptyset \forall s \neq t,$

Minimizing (1) partitions the domain $\Omega \subset \mathbb{R}^d$ into $k + 1$ pairwise disjoint sets E_l . The first term measures the perimeter of the set E_l which leads to smooth segmentation boundaries. The second term is the data term which is based on non-negative weight functions $f_l : \Omega \rightarrow \mathbb{R}^+$. Model (1) includes as a special case the piecewise constant Mumford-Shah functional [24], which arises when choosing

$$f_l(x) = \lambda(I(x) - c_l)^2, \quad (2)$$

which is the squared difference of the input image $I(x)$ to some mean intensity c_l scaled by λ (cf. Figure 1).

Several algorithms have been proposed to approximately minimize the Potts model. While the discrete problem can be tackled using iterated binary optimization via α -expansion [5] or roof duality relaxation [27], such algorithms tend to exhibit a grid bias (metrication errors) in representing the continuous perimeters in (1). In the continuous domain, popular methods are based on the level set method (e.g. Chan and Vese [13]) The most crucial drawback of these methods is, that there is no guarantee to find a globally optimal solution.

In this paper, we propose a convex relaxation of the Potts model. It is formulated in a continuous setting and therefore does not exhibit metrication errors. The proposed method also allows to compute a bound between the solution of the relaxed problem and the true optimum. In practice, we often observe convergence to globally optimal solutions.

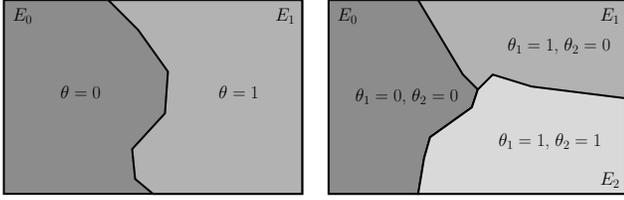
2. Convex Representation of the Potts Model

In this section we develop a convex representation of the Potts model. We start by reviewing the two-label case which was presented in [11]. We then show how to generalize their work to the multi-label case.

2.1. The Two-Label Case: $k = 1$

In [11], Chan, Esedoglu and Nikolova gave a convex formulation of (1) for $k = 1$, by rewriting it in terms of the variational model

$$\min_{\theta} \left\{ \int_{\Omega} |D\theta| + \int_{\Omega} (1 - \theta(x))f_0(x) + \theta(x)f_1(x) dx \right\}, \quad (3)$$



(a) Two label case (b) Three label case

Figure 2. (a) The two-label case: A binary function θ is used to partition the image domain Ω into two regions. (b) The three-label case: Two binary functions $\theta_1 \geq \theta_2$ are used to partition the image domain Ω into three regions.

where $\theta : \Omega \rightarrow \{0, 1\}$ is a binary function, used to model the partition of the image domain into two regions, i.e. $\theta(x) = 0$ if $x \in E_0$ and $\theta(x) = 1$ if $x \in E_1$. See Figure 2(a), where the function θ is used to partition the image domain Ω into two regions. The term $\int_{\Omega} |D\theta|$ is the Total Variation of θ . For binary functions θ , it is equal to the total interface area. $D\theta$ is the distributional derivative which is, in an integral sense, also well-defined for non-smooth functions θ . On the other hand, if θ is a smooth function, the Total Variation of θ is simply $\int_{\Omega} |\nabla\theta| dx$.

There exists also a more general formulation of the Total Variation [7, 12, 9]

$$\int_{\Omega} |D\theta| = \sup_{\xi: |\xi(x)| \leq 1} \left\{ - \int_{\Omega} \theta \operatorname{div} \xi \, dx \right\}, \quad (4)$$

where $\xi = (\xi^1, \dots, \xi^d)^T : \Omega \rightarrow \mathbb{R}^d$ is the dual variable and $|\cdot|$ denotes the Euclidean vector norm. This formulation is also called the dual formulation of the Total Variation. The main advantage of the dual formulation over the original formulation is, that it is valid for any L^1 integrable function, but it comes along with inequality constraints on the dual variable ξ which brings some additional complexity to the formulation.

2.1.1 Convex Relaxation

Although (3) is convex in θ , the variational problem is still non-convex since the minimization is carried out over binary functions θ , which comprise a non-convex set of functions. A straightforward approach is to relax (3) by allowing θ to vary smoothly in the interval $[0, 1]$. Furthermore, it can be shown that solutions of the relaxed problem can be transformed to solutions of the binary problem by thresholding of the solution of the relaxed problem. It is remarkable that this works for any threshold $s \in (0, 1)$.

2.2. The General Case: $k \geq 1$

Let $u : \Omega \rightarrow \{0, \dots, k\}$ be the labeling function such that $u(x) = l$ if and only if $x \in E_l$. We can equivalently

represent this multi-label function by k binary functions $\boldsymbol{\theta}(x) = (\theta_1(x), \dots, \theta_k(x))$ defined by

$$\theta_l(x) = \begin{cases} 1 & \text{if } u(x) \geq l \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

representing its upper level sets. In turn, the labeling function u can be recovered from these functions via the relation

$$u(x) = \sum_{l=1}^k \theta_l(x). \quad (6)$$

Figure 2(b) shows an example for $k = 2$, where two functions θ_1 and θ_2 are used to partition the image domain Ω into three regions.

A one-to-one correspondence between multi-label functions $u(x)$ and vectors $\boldsymbol{\theta}(x)$ of binary functions is guaranteed by imposing the ordering

$$0 \leq \theta_k(x) \leq \dots \leq \theta_1(x) \leq 1, \quad \forall x \in \Omega. \quad (7)$$

It is therefore appropriate, to define the set of admissible vector functions $\boldsymbol{\theta}$ as

$$\mathcal{B}_0 = \left\{ \boldsymbol{\theta} = (\theta_1, \dots, \theta_k) : \Omega \rightarrow \{0, 1\}^k, \right. \\ \left. 0 \leq \theta_k(x) \leq \dots \leq \theta_1(x) \leq 1, \quad \forall x \in \Omega \right\}. \quad (8)$$

Furthermore, from the definition of (5) one can see that we can use the following relation to indicate whether a pixel belongs to the set E_l or not:

$$\theta_l(x) - \theta_{l+1}(x) = \begin{cases} 1 & \text{if } u(x) = l \\ 0 & \text{otherwise} \end{cases}, \quad (9)$$

where we use $\theta_0(x) = 1$ and $\theta_{k+1} = 0$ to simplify notations. The goal is now to rewrite the Potts model (1) in terms of $\boldsymbol{\theta}$. First, we can make use of (9) to rewrite the data term of the Potts model as

$$\sum_{l=0}^k \int_{E_l} f_l(x) \, dx = \sum_{l=0}^k \int_{\Omega} (\theta_{l+1}(x) - \theta_l(x)) f_l(x) \, dx \quad (10)$$

The length term in (1) requires some additional considerations. A natural generalization of (3) would be to simply approximate the boundary length by the length of the binary sets:

$$\sum_{l=0}^k \operatorname{Per}(E_l; \Omega) \approx \sum_{l=0}^k \int_{\Omega} |D\theta_l|. \quad (11)$$

Unfortunately, this turns out to be a bad approximation. As shown in Figure 2(b), interfaces where m functions θ_l jump from 0 to 1 are counted m times. As a result, the approximation (11) would count the boundary between E_0 and E_2 twice.

How can this systematic error be resolved in a computationally tractable manner? It turns out that the dual formulation (4) allows to suppress a multiple counting of boundaries when combined with an additional constraint on the dual variables. More specifically we have the equality:

$$\sum_{l=0}^k \text{Per}(E_l; \Omega) = \sup_{\xi \in \mathcal{K}} \left\{ \sum_{l=0}^k - \int_{\Omega} \theta_l \text{div } \xi_l \right\}, \quad (12)$$

where the dual variables $\xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^{d \times k}$ are constrained to lie in the set \mathcal{K} defined as:

$$\mathcal{K} = \left\{ \xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^{d \times k}, \right. \\ \left. \left| \sum_{l_1 \leq l \leq l_2} \xi_l(x) \right| \leq 1, \forall x \in \Omega, 1 \leq l_1 \leq l_2 \leq k \right\}.$$

For the case shown in Figure 2(b), for example, the above constraint implies that $|\xi_1(x) + \xi_2(x)| \leq 1$. This assures that the transition between E_0 and E_2 is counted exactly once.

The following proposition is of central importance to assure a computationally tractable algorithm:

Proposition 1. *The set \mathcal{K} is convex.*

Proof. Let $\xi, \xi' \in \mathcal{K}$. Using the triangle inequality it follows for any $\alpha \in [0, 1]$ that $\alpha\xi + (1 - \alpha)\xi' \in \mathcal{K}$:

$$\left| \sum_{l_1 \leq l \leq l_2} \alpha \xi_l(x) + (1 - \alpha) \xi'_l(x) \right| \leq \\ \alpha \left| \sum_{l_1 \leq l \leq l_2} \xi_l(x) \right| + (1 - \alpha) \left| \sum_{l_1 \leq l \leq l_2} \xi'_l(x) \right| \leq 1, \quad \square$$

We are now ready to give a convex representation of the Potts model.

$$\min_{\theta \in \mathcal{B}_0} \sup_{\xi \in \mathcal{K}} \left\{ \sum_{l=0}^k - \int_{\Omega} \theta_l \text{div } \xi_l dx \right. \\ \left. + \int_{\Omega} (\theta_{l+1}(x) - \theta_l(x)) f_l(x) dx \right\}. \quad (13)$$

Note that θ still has to fulfill the ordering constraints of (7). To further simplify the problem, it turns out that we can eliminate the ordering constraints by slightly modifying the energy functional to

$$\mathcal{E} = \min_{\theta \in \mathcal{B}} \sup_{\xi \in \mathcal{K}} \left\{ \sum_{l=0}^k - \int_{\Omega} \theta_l \text{div } \xi_l dx \right. \\ \left. + \int_{\Omega} |\theta_{l+1}(x) - \theta_l(x)| f_l(x) dx \right\}, \quad (14)$$

where the convex set \mathcal{B} is defined as

$$\mathcal{B} = \left\{ \theta = (\theta_1, \dots, \theta_k) : \Omega \rightarrow \{0, 1\}^k \right\}. \quad (15)$$

The equivalence of minimizers is characterized by the following proposition:

Proposition 2. *The minimizer of problem (13) is equivalent to the minimizer of problem (14).*

Proof. To proof the equivalence of the minimizers it suffices to show that the minimizer of (14) is ordered in the sense of (7). A complete proof is presented in [10], Proposition 4.3. \square

2.2.1 Convex Relaxation

In order to solve (14) we apply the same convexification as in the two-label case. We again relax the set of binary solutions to the set of functions which can take all values in $[0, 1]$, that is

$$\mathcal{R} = \left\{ \theta = (\theta_1, \dots, \theta_k) : \Omega \rightarrow [0, 1]^k \right\}. \quad (16)$$

Unfortunately, the thresholding technique of the two-label case no longer holds in the general setting. However, we can give a bound on the energy difference between the true minimizer of the binary problem and the minimizer of the relaxed problem.

Proposition 3. *Let $\theta^* \in \mathcal{R}$ be the solution of the relaxed problem (14) and let $\mathbf{1}_{\{\theta^* \geq s\}} \in \mathcal{B}$ be a thresholded binary version for any $s \in [0, 1]$. Furthermore, let $\theta' \in \mathcal{B}$ be the true global minimizer of the binary problem (14). Then we can provide the following bound on the solution of the relaxed problem.*

$$|\mathcal{E}(\theta^*) - \mathcal{E}(\theta')| \leq |\mathcal{E}(\theta^*) - \mathcal{E}(\mathbf{1}_{\{\theta^* \geq s\}})|. \quad (17)$$

Proof. It suffices to show that energetically the thresholded solution lies between the optimal of the relaxed problem and the optimal of the binary problem. First, by convexity of (14), the energy of the relaxed solution is the lowest. Second, any thresholded version, i.e. a projection of the relaxed solution onto the binary set must achieve a higher energy: $\mathcal{E}(\theta^*) \leq \mathcal{E}(\mathbf{1}_{\{\theta^* \geq s\}})$. Clearly the energy of the true binary minimizer lies somewhere between the energy of the relaxed solution and the energy of the thresholded version, i.e. $\mathcal{E}(\theta^*) \leq \mathcal{E}(\theta') \leq \mathcal{E}(\mathbf{1}_{\{\theta^* \geq s\}})$. \square

As a direct consequence of Proposition 3 we see that, if the solution of the relaxed problem is already binary, then we have found a true global minimizer of the binary problem, i.e. we have found the solution of the Potts model.

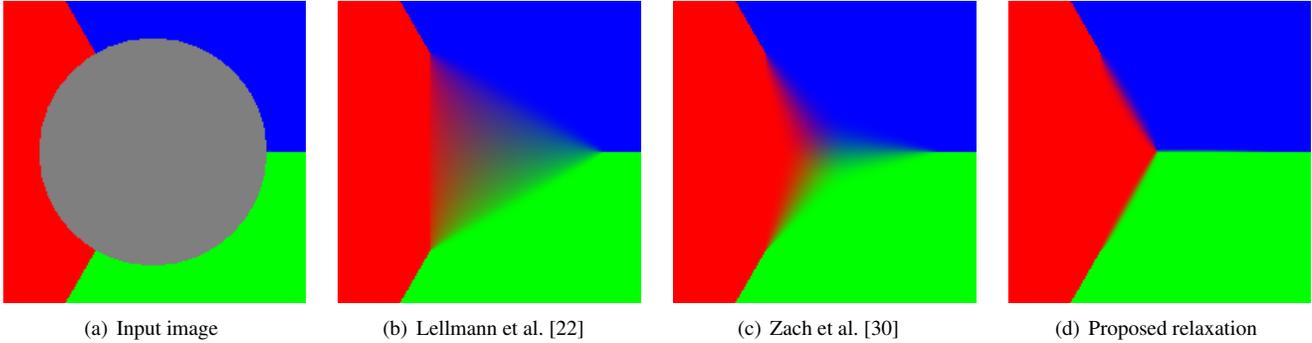


Figure 3. Synthetic three-label image segmentation problem. The optimal segmentation is given by a 120° triple junction. The alternative relaxation techniques (b) and (c) end up with a non-binary solutions whereas our method (d) finds an almost binary solution.

2.2.2 Alternative Approaches

Zach et al. in [30] proposed a different relaxation approaches to tackle the Potts model. More recently, Lellman et al. in [22] proposed a qualitatively comparable approach with a slightly different length term. The idea is to introduce a relaxed labeling function $\mathbf{v} = (v_0, \dots, v_k)$, with $v_l(x) \in [0, 1]$ and the equality constraint $\sum_{l=0}^k v_l(x) = 1$. Then the Potts model (1) simply becomes

$$\min_{\mathbf{v}} \left\{ \frac{1}{2} \sum_{l=0}^k \int_{\Omega} |Dv_l| + \sum_{l=0}^k \int_{\Omega} v_l(x) f_l(x) dx \right\}. \quad (18)$$

Indeed, for $v_l(x) \in \{0, 1\}$, this energy clearly equals the Potts energy (1). Yet, the relaxation is not as tight as the one we have proposed in (14).

Proposition 4. *The relaxation (14) strictly dominates (18).*

Proof. It suffices to show that the set \mathcal{K} used in (14) allows for a larger set of feasible functions than the above approach. For space limitations we refer to the technical report [10] for more details. \square

Figure 3 shows a comparison of the proposed approach to the relaxation approach of [30] and [22] by means of a simple three-label segmentation problem. We used the data term (2), where the means c_l correspond to pure red $(1, 0, 0)$ green $(0, 1, 0)$ and blue $(0, 0, 1)$. The input image also contains a grey area, where all data terms are equal, i.e. the length term dominates in this region. It is remarkable that the proposed method finds the true binary solution, a triple junction where all three interfaces meet with an angle of 120° . On the other hand the alternative approaches do not find a binary solution. This is also reflected by the obtained energies. The solution of our approach achieves an energy of $\mathcal{E} = 13975$ which is clearly above $\mathcal{E} = 13972$ achieved by [30] and $\mathcal{E} = 13960$ achieved by [22]. The energy of the true binary segmentation is $\mathcal{E} = 14006$. Note that a higher energy also gives a tighter bound in (17).

3. Numerical Algorithms

In this section we present numerical algorithms to compute the solution of the relaxed problem (14).

3.1. Discrete Setting

We consider only the 2D case ($d = 2$) here we assume that $\Omega = (0, 1)^2$. For discretization, we use a standard Cartesian grid defined by

$$\Omega^h = \{(ih, jh) : 1 \leq i, j \leq N\}, \quad (19)$$

where (i, j) are the indices of the discrete locations on the grid and the width of the spatial discretization is given by $h = 1/N$. In the following we will use the superscript “h” to indicate the discrete setting. A spatially continuous function u is now approximated by a spatially discretized function $\theta^h : \Omega^h \rightarrow \mathbb{R}$. Next, we define a discrete version of the Total Variation

$$h^2 \sum_{i,j} |(\nabla \theta^h)_{i,j}| = \max_{|\xi_{i,j}^h| \leq 1} -h^2 \sum_{i,j} \theta_{i,j}^h (\text{div } \xi^h)_{i,j}, \quad (20)$$

where $|\cdot|$ denotes the standard Euclidean norm, ∇ and div are now discrete versions of the gradient and divergence operators involving a $1/h$ factor. See [9] for more details. Before we proceed by defining a discrete version of (14), let us give the following fundamental result which applies for our type of discretization.

Proposition 5. *The discrete version of the Total Variation Γ -converges to the Total Variation as $h \rightarrow 0$.*

The complete proof of this proposition is presented in [10], Proposition 3.1. This proposition states that by only refining the discretization width h , the discrete Total Variation converges to the true continuous Total Variation. Note that this is not the case for MRF based approaches, where in addition to the grid refinement, the number of pairwise interactions has to go to infinity [4]. Note that this proposition is also true for the Potts model.

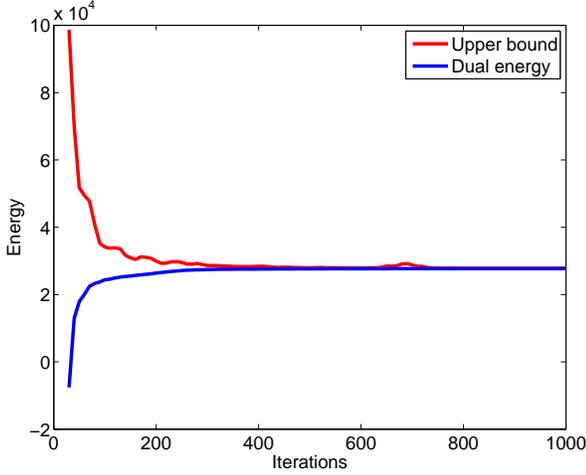


Figure 4. Convergence of the primal dual algorithm for the Tsukuba stereo example of Figure 7. The upper bound can be used to determine convergence of the algorithm.

3.2. Primal-Dual Algorithm

In order to make (14) continuously differentiable, we introduce an additional dual variable $\boldsymbol{\eta} = (\eta_1, \dots, \eta_k) : \Omega \rightarrow \mathbb{R}^{d \times k}$ such that, in a discrete setting (14) can be rewritten as

$$\min_{\boldsymbol{\theta}^h \in \mathcal{R}^h} \max_{\substack{\boldsymbol{\xi}^h \in \mathcal{K}^h \\ \boldsymbol{\eta}^h \in \mathcal{W}^h}} \left\{ \sum_{l=0}^k -h^2 \sum_{i,j} (\theta_l^h)_{i,j} (\text{div } \xi_l^h)_{i,j} + h^2 \sum_{i,j} ((\theta_{l+1}^h)_{i,j} - (\theta_l^h)_{i,j}) (\eta_l^h)_{i,j} \right\}. \quad (21)$$

We have also introduced the convex set \mathcal{W}^h is given by

$$\mathcal{W}^h = \left\{ \boldsymbol{\eta}^h = (\eta_1^h, \dots, \eta_k^h) : \Omega^h \rightarrow \mathbb{R}^k, \right. \\ \left. |(\eta_l^h)_{i,j}| \leq (f_l^h)_{i,j}, \forall (i,j) \in \Omega^h, 1 \leq l \leq k \right\}. \quad (22)$$

The algorithm we propose for minimizing (21) is an Arrow-Hurwicz type primal-dual gradient flow algorithm [2]. Basically the algorithm consists of alternating a gradient descend in the primal variable and a gradient ascend in the dual variable. After each update step the primal and dual variables are re-projected to the respective sets. The outline of the algorithm is as follows:

1. We fix the time steps $\tau_p > 0$ and $\tau_d > 0$.
2. Let n be the iteration number. We set $n = 0$ and initialize $(\boldsymbol{\theta}^h)^0 = \mathbf{0}$, $(\boldsymbol{\xi}^h)^0 = \mathbf{0}$ and $(\boldsymbol{\eta}^h)^0 = \mathbf{0}$.
3. For all i, j, l update the primal and dual variables by

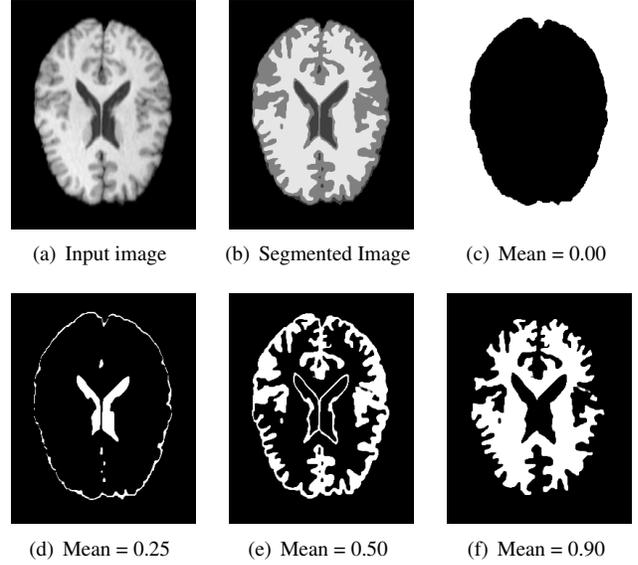


Figure 5. Gray and white matter segmentation of a brain MRI image with 4 labels, fixed means and using $\lambda = 50$. (a) shows the input image, (b) shows the piecewise constant approximation and (c)-(f) depict the corresponding labels.

the following projected gradient schemes:

$$\begin{aligned} (\theta_l^h)_{i,j}^{n+1} &= \Pi_{\mathcal{R}^h} \left((\theta_l^h)_{i,j}^n + \tau_p (\text{div } \xi_l^h)_{i,j} \right), \\ (\xi_l^h)_{i,j}^{n+1} &= \Pi_{\mathcal{K}^h} \left((\xi_l^h)_{i,j}^n + \tau_d (\nabla \theta_l^h)_{i,j} \right), \\ (\eta_l^h)_{i,j}^{n+1} &= \Pi_{\mathcal{W}^h} \left((\eta_l^h)_{i,j}^n + \tau_d (\theta_{l+1}^h - \theta_l^h)_{i,j} \right), \end{aligned}$$

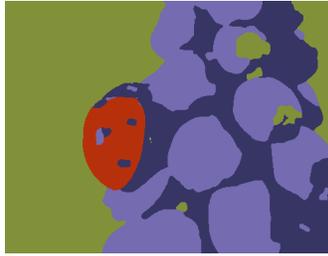
where $\Pi_{\mathcal{R}^h}$, $\Pi_{\mathcal{K}^h}$ and $\Pi_{\mathcal{W}^h}$ are Euclidean projectors to the respective convex sets.

We found the numerical scheme to be stable as long as the product $\tau_p \tau_d \leq 1/(3h^2)$. This bound seems to be sharp, we do not have any proof for that. In practice we choose $\tau_p = \tau_d = 1/(\sqrt{3}h)$. In our numerical scheme we need to re-project the primal and dual variables on the convex sets \mathcal{R}^h , \mathcal{K}^h and \mathcal{W}^h . The projections of $\boldsymbol{\theta}^h$ onto \mathcal{R}^h and $\boldsymbol{\eta}^h$ onto \mathcal{W}^h are easy. They can be performed by simple point-wise truncation operations. The projection of $\boldsymbol{\xi}^h$ onto \mathcal{K}^h is more complicated since it involves constraints over several levels. In order to perform the projection, we use the iterative projection algorithm of Dykstra [6].

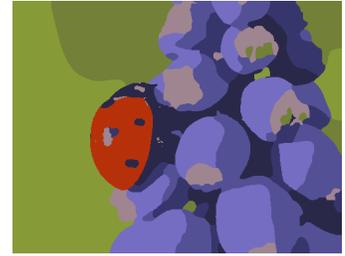
Our primal-dual algorithm allows for computing the so-called primal-dual gap $\mathcal{G} = \mathcal{E}_p - \mathcal{E}_d$, where \mathcal{E}_p is the maximum of (21) with respect to the dual variables $(\boldsymbol{\xi}^h, \boldsymbol{\eta}^h)$ and \mathcal{E}_d is the minimum of (21) with respect to the primal variable $\boldsymbol{\theta}^h$. Computing the energy of the dual problem \mathcal{E}_d is easy since it can be computed by solving simple point-wise optimization problems. However, computing the energy of the primal problem \mathcal{E}_p can be tedious since it will itself need an iterative algorithm to compute the maximum



(a) Input image



(b) Segmentation for $k = 4, \lambda = 4$

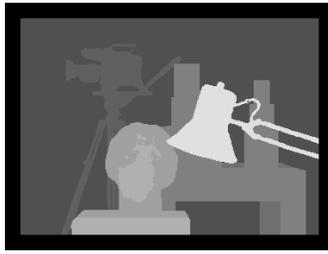


(c) Segmentation for $k = 8, \lambda = 5$

Figure 6. Piecewise constant Mumford-Shah color segmentation of a natural image using a different number of labels.



(a) Left input image



(b) Ground truth disparity



(c) Estimated disparity map with occlusions

Figure 7. Stereo estimation using 16 labels. Occluded pixels were determined using a left-right consistency check.

with respect to the dual variables. We therefore propose to compute an upper bound of the primal energy by computing the Potts energy of a thresholded version of the primal variable. Note that in this case the gap is not guaranteed to go to zero but it still provides a suitable criterion to check for convergence. Figure 4 shows the dual energy and the corresponding upper bound on the primal energy computed for the stereo example provided in Figure 7. One can see that after 1000 iterations, the dual energy is very close to the upper bound of the primal energy.

3.3. Parallel Implementation

The first order primal dual algorithm we developed in the last section is known to be slow in terms of the number of iterations, but it turned out to be convenient, since it can be effectively parallelized on GPUs. We therefore implemented the algorithm using the CUDA framework of NVidia. Using a Nvidia Tesla C1060 GPU, we achieved a speedup of approximately a factor 30 compared to an optimized C/C++ implementation. Typical runtimes for image segmentation problems for 512^2 images and 10 labels are in the range of 50 seconds.

4. Experimental Results

In the first experiment, we apply our algorithm to gray and white matter segmentation of a brain MRI image which is a classical problem in medical image analysis. Figure 5 shows the segmentation result using the piecewise constant

Mumford-Shah model for 4 labels and using the data term (2) using $\lambda = 50$. The mean values were hand selected. One can see that the functions $\theta_l - \theta_{l+1}$ are binary, i.e. the algorithm has found the optimal labeling.

In the second experiment we apply the algorithm to color segmentation of a natural image using again the data term in (2). In this case the mean values were initialized using a k-means algorithm. After computing the optimal segmentation, the mean values were updated accordingly and the segmentation process was repeated. Figure 6 shows the segmentation result using 4 and 8 labels. The parameter λ was set to 4 and 5, respectively

In the last experiment, we apply the proposed method to disparity estimation on a standard benchmark stereo image pair. We use 16 labels corresponding to disparity values ranging from $0 \dots 15$. The data term was the pixel-wise sum of the absolute differences of the RGB color channels. Figure 7 depicts the left input image, the ground truth disparity and the compute disparity. Occluded pixels were determined using a left-right consistency check. Note that the algorithm accurately restores the disparity of fine structures such as the lamp.

5. Conclusion

We proposed a convex relaxation approach for minimizing the spatially continuous Potts model. Based on a primal dual formulation of the Total Variation functional we showed that the Potts prior can be elegantly incorporated by

means of convex constraints on the dual variables. For minimization we propose an efficient primal dual projected gradient algorithm which also allows a fast implementation on parallel hardware. Although our approach is not guaranteed to find global minimizers of the Potts model we can give a tight bound on the energy between the computed solution and the true minimizer. Furthermore we show that our relaxation approach dominates recently proposed relaxations. This implies that computed solutions are closer to the true minimizer. For many practical problems we actually find global minimizers. We demonstrate the excellent performance of our approach on problems of multi-label image segmentation and stereo reconstruction.

References

- [1] G. Alberti, G. Bouchitté, and D. Dal Maso. The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calc. Var. Partial Differential Equations*, 16(3):299–333, 2003.
- [2] K. J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in linear and non-linear programming*. With contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow. Stanford Mathematical Studies in the Social Sciences, vol. II. Stanford University Press, 1958.
- [3] D. Bertismas and J. Tsitsikilis. *Introduction to linear optimization*. Athena Scientific, 1997.
- [4] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Proc. 9th Int. Conf. Comp. Vis.*, pages 26–33, Nice, France, 2003.
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001.
- [6] J. P. Boyle and R. Dykstra. A method for finding projections onto the intersection of convex sets in Hilbert spaces. In *Advances in order restricted statistical inference (Iowa City, Iowa, 1985)*, volume 37 of *Lecture Notes in Statist.*, pages 28–47. Springer, Berlin, 1986.
- [7] J. Carter. *Dual Methods for Total Variation-based Image Restoration*. PhD thesis, UCLA, Los Angeles, CA, 2001.
- [8] A. Chambolle. Convex representation for lower semicontinuous envelopes of functionals in L^1 . *J. Convex Anal.*, 1(8):149–170, 2001.
- [9] A. Chambolle. An algorithm for total variation minimizations and applications. *J. Math. Imaging Vis.*, 2004.
- [10] A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Technical Report CMAP-649, Ecole Polytechnique, Paris, France, 2008.
- [11] T. Chan, S. Esedoglu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal of Applied Mathematics*, 66(5):1632–1648, 2006.
- [12] T. Chan, G. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comp.*, 20(6):1964–1977, 1999.
- [13] T. Chan and L. Vese. Active contours without edges. *IEEE Trans. Image Process.*, 10(2):266–277, February 2001.
- [14] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *J. Royal Statistics Soc.*, 51(Series B):271–279, 1989.
- [15] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13, 1902.
- [16] P. L. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 0-1 optimization. *Math. Programming*, 28(2):121–155, 1984.
- [17] H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(10):1333–1336, 2003.
- [18] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 23:253–258, 1925.
- [19] R. Kindermann and J. L. Snell. *Markov Random Fields and Their Applications*. American Mathematical Society, 1980.
- [20] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *European Conference on Computer Vision (ECCV)*, 2008.
- [21] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(2):147–159, 2004.
- [22] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. Technical report, IWR, University of Heidelberg, 2008.
- [23] D. Mumford and A. Desolneux. *Pattern Theory: The Stochastic Analysis of Real World Signals*. in progress, 2008.
- [24] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
- [25] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *European Conference on Computer Vision (ECCV)*, 2008.
- [26] R. B. Potts. Some generalized order-disorder transformations. *Proc. Camb. Phil. Soc.*, 48:106–109, 1952.
- [27] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *Proc. IEEE Conf. Comp. Vis. Pattern Recognit.*, pages 1–8, 2007.
- [28] D. Schlesinger and B. Flach. Transforming an arbitrary min-sum problem into a binary one. Technical Report TUD-FI06-01, Dresden University of Technology, 2006.
- [29] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *IEEE Trans. Inf. Theory*, 51(11):3697–3717, 2005.
- [30] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, Modeling and Visualization Workshop (VMV)*, 2008.