

A Survey and Comparison of Discrete and Continuous Multi-label Optimization Approaches for the Potts Model

Claudia Nieuwenhuis · Eno Töppe · Daniel Cremers

Received: 16 December 2011 / Accepted: 27 February 2013
© Springer Science+Business Media New York 2013

Abstract We present a survey and a comparison of a variety of algorithms that have been proposed over the years to minimize multi-label optimization problems based on the Potts model. Discrete approaches based on Markov Random Fields as well as continuous optimization approaches based on partial differential equations can be applied to the task. In contrast to the case of binary labeling, the multi-label problem is known to be NP hard and thus one can only expect near-optimal solutions. In this paper, we carry out a theoretical comparison and an experimental analysis of existing approaches with respect to accuracy, optimality and runtime, aimed at bringing out the advantages and short-comings of the respective algorithms. Systematic quantitative comparison is done on the Graz interactive image segmentation benchmark. This paper thereby generalizes a previous experimental comparison (Klodt et al. 2008) from the binary to the multi-label case.

Keywords Multi-label · Survey · Comparison · Optimization · Markov random fields · Partial differential equations

1 Introduction

The optimization of energies with respect to a set of variables which can take one of multiple labels is among the central algorithmic challenges in computer vision and image analysis. The prototypical example of multi-label problems is multiregion image segmentation, where every pixel is assigned

one of finitely many region labels. Apart from segmentation, continuous estimation problems such as denoising, deblurring, stereo and optical flow can be approximated as a multi-label problem on a discretized label space. We will restrict our attention to algorithms which aim at minimizing a specific class of multiregion segmentation functionals, often referred to as the *Potts problem* in the MRF community or the *minimal partition problem* in the PDE community.

Over the years, numerous algorithms have been proposed to tackle multi-label optimization problems, both in the community of *partial differential equations (PDEs)* and in the community of *Markov random fields (MRFs)*. As a result one may ask how these algorithms compare in theory and in practice. Klodt et al. (2008) presented an experimental comparison of discrete and continuous optimization approaches for the specific case of *binary* labeling problems. In recent years, the focus has shifted from binary labeling to the more general multi-label problem, with a multitude of competing algorithms to solve it. The contribution of this paper is to provide a systematic theoretical and experimental comparison of algorithms for multi-label problems of the minimal partition type, pointing out relations, equivalences and differences.

In general, the segmentation task can be formulated as an energy minimization problem. In the spatially discrete setting, this energy is defined on a set of nodes, leading to MRF problems whose solution is often calculated using graph cut methods. In the spatially continuous setting, the respective optimality conditions for the continuous energy are written in terms of a set of partial differential equations which are then solved on a discrete grid.

The contributions of this paper are twofold: Firstly, we present relations between the following relaxations of multi-label optimization problems from the MRF and PDE communities:

C. Nieuwenhuis (✉) · E. Töppe · D. Cremers
Departments of Computer Science and Mathematics,
TU München, Germany
e-mail: claudia.nieuwenhuis@in.tum.de

- The linear KT-relaxation (Kleinberg and Tardos (2002)),
- The linear S-relaxation (Schlesinger (1976), revisited by Werner (2007)).
- The convex CCP-relaxation (Chambolle et al. (2008)),
- The convex ZGFN-relaxation (Zach et al. (2008)),

where the acronyms denote the initials of the authors' last names. Secondly, we qualitatively and quantitatively compare the following popular PDE- and graph cut-based algorithms for multi-label segmentation from the continuous and discrete domain:

- The primal–dual algorithm (Pock et al. (2009)),
- The discrete α -expansion approach with four and eight connectivity (Boykov et al. 2001),
- The discrete Fast-PD approach with four and eight connectivity (Komodakis et al. 2007).

A preliminary version of this work was published in Nieuwenhuis et al. (2011). We plan to make code publically available at '<https://vision.in.tum.de/data/software>'.

2 Advantages and Drawbacks of Discrete and Continuous Methods

Concerning graph cut based MRF approaches and PDE methods we encountered several advantages and short-comings. In particular:

2.1 Parameters and Termination Criterion

Graph cuts do not require numerical tuning parameters, so they can easily be applied as a blackbox algorithm without further need to understand the underlying optimization process. Instead, PDE approaches typically require more expert knowledge. Optimal step sizes can be computed automatically (Pock and Chambolle (2011)). However, general-purpose termination criteria are hard to find. Typically one performs a fixed number of optimization cycles or iterates until the change in the solution or in the energy falls below a certain threshold. Alternatively one can threshold the primal–dual gap, i.e. the difference between the primal and dual energies which goes to zero upon convergence. Depending on the problem formulation, this gap can either be computed by solving a simple point-wise optimization problem or it can be as complicated to determine as the original problem. In the latter case, only upper or lower energy bounds can be computed leading to approximations of the primal–dual gap but no hard termination criteria.

2.2 Ambiguities

Since we can only compute approximate solutions to the multi-label segmentation problem, results come with certain ambiguities. The commonly used α -expansion and Fast-PD algorithms are based on iteratively solving binary problems. In each step, each pixel is allowed to choose between its current label and a fixed label α . As we show in Fig. 5, results may therefore depend on the order the labels α are chosen and also on the initialization. In the PDE domain ambiguities arise *after* the optimization stage, when the globally optimal solution of the relaxed problem is binarized to obtain a feasible labeling of the original problem. The results, however, do not depend on the initialization of the algorithm due to the convexity of the relaxations.

2.3 Metrication Errors

MRF approaches exhibit metrication errors since they only approximate the Euclidean boundary length—see Sect. 7.1.2. Instead, the PDE approaches for multi-label optimization provide smooth object boundaries that do not exhibit a prominent grid bias.

2.4 Parallelization Potential

Graph cuts cannot be parallelized in a straight forward manner, since the max-flow problem lies in the P-complete complexity class of problems, which are probably not efficiently parallelizable (Goldschlager et al. 1982). The popular *augmenting path* algorithms sequentially search for paths through the flow network along which the flow can be incremented. Regardless of the various strategies that exist to find such a path, these algorithms are not well suited for parallelization due to two reasons: Firstly, augmenting path operations are interdependent, since different augmentation paths can share edges. Secondly, the updates of the edge residuals have to be carried out simultaneously in each augmentation operation as they all depend on the minimum capacity within the augmentation path. Although *push relabel* algorithms relax the first issue, update operations are still interdependent. In contrast to MRF approaches, the considered PDE approaches are easily parallelizable on graphics hardware yielding drastic speedups over CPU algorithms. However, they come with the drawback of an increased power consumption per pixel.

2.5 Runtime Variance

Graph cut methods exhibit a high runtime variance even among images with the same number of labels (see Table 4).

There are mainly two reasons for this: firstly, for α -expansion the number of max flow problems that need to be solved until convergence highly depends on the input image and the chosen label order. Secondly, the number of augmentation steps needed to solve each of the max flow problems in turn depends on the graph structure and edge residuals. These parameters strongly differ with the current labeling and the currently chosen label α . In contrast, PDE approaches carry out the same computation steps on each pixel and thus exhibit smaller runtime variances. Nevertheless, the number of iterations until convergence depends on the input image.

While the above advantages and drawbacks are more or less known today, systematic practical comparisons of both worlds concerning quality and runtimes for multi-label partitioning problems have not been undertaken so far. Hence, in this paper we extend the research by Klodt et al. (2008) to multi-label problems and investigate theoretical and practical differences for the most prevalent relaxations and algorithms from the MRF and PDE community.

3 Image Segmentation as a Multi-label Problem

In the context of image segmentation, the multi-label problem can be formulated as follows: let $I : \Omega \rightarrow \mathbb{R}^d$ denote the input image defined on the domain $\Omega \subset \mathbb{R}^2$ (in the discrete case $\Omega = \{1, \dots, N\} \times \{1, \dots, M\}$). The task of segmenting the image plane into a set of n pairwise disjoint regions Ω_i , with $\Omega = \bigcup_{i=1}^n \Omega_i$, $\Omega_i \cap \Omega_j = \emptyset \quad \forall i \neq j$, can be solved by computing a labeling $l : \Omega \rightarrow \{1, \dots, n\}$ indicating which of the n regions each pixel belongs to: $\Omega_i = \{x \mid l(x) = i\}$.

Segmentation can be formulated as an energy minimization problem consisting of a data term and a regularization term. The data term carries information on how strongly a pixel should be associated with each region. The regularization term corresponds to a prior on the space of feasible solutions. The most common regularizer for segmentation is the one favoring minimal boundary length. We will give the corresponding energy minimization problems in the spatially continuous domain in Sect. 4 and in the spatially discrete domain in Sect. 5.

4 The Spatially Continuous Setting

In this section, we focus on energy minimization problems for multi-label segmentation which are solved by means of variational calculus. We will review the two convex relaxations by Chambolle et al. (2008) and Zach et al. (2008), which allow for near-optimal solutions to the multi-label segmentation problem.

4.1 Minimal Partitions and Mumford-Shah

The multi-label segmentation problems we consider are of the following form

$$\min_{\Omega_1, \dots, \Omega_n} \left\{ \frac{\lambda}{2} \sum_{i=1}^n \text{Per}_g(\Omega_i, \Omega) + \sum_{i=1}^n \int_{\Omega_i} f_i(x) dx \right\}. \quad (1)$$

Here, $\text{Per}_g(\Omega_i, \Omega)$ denotes the perimeter of the segment Ω_i , measured with a metric defined by the non-negative function $g : \Omega \rightarrow \mathbb{R}_0^+$. For example, the choice $g(x) = \exp(-\gamma|\nabla I(x)|)$ energetically favors boundaries that coincide with strong gradients of the input image $I : \Omega \rightarrow \mathbb{R}$. The data term $f_i : \Omega \rightarrow \mathbb{R}$ associated with region Ω_i takes on smaller values if the respective pixel (based on its color) has stronger affinity to region Ω_i . In a Bayesian MAP inference approach, $f_i = -\log P(I, x | \Omega_i)$ corresponds to the negative logarithm of the conditional probability for observing a specific color at the given location within region Ω_i —see also Cremers et al. (2007) and Nieuwenhuis and Cremers (2012).

In the continuous setting, the minimal partition problem is related to the piecewise constant case of the Mumford-Shah model (Mumford and Shah 1989) which aims at segmenting the image plane into a set of pairwise disjoint regions with minimal color variance and minimal boundary length. Optimization of this functional is difficult. Among the most popular techniques are local optimization approaches such as level set methods (Chan and Vese 2001; Tsai et al. 2001).

4.2 Total Variation

More recently, minimal partition problems have been tackled on the basis of functions of bounded variation $BV(\Omega, \mathbb{R})$, i.e. functions for which the total variation (TV) is finite. The key idea is to encode the regions Ω_i by their indicator function $u \in BV(\Omega, \{0, 1\})^n$

$$u_i(x) = \begin{cases} 1, & \text{if } l(x) = i \\ 0, & \text{otherwise} \end{cases} \quad \forall i = 1, \dots, n \quad (2)$$

and to solve for u_i using convex relaxation techniques.

The *total variation* is well suited for geometric optimization problems since it is convex and—for binary functions u_i —equal to the perimeter of the encoded set Ω_i . Let $\xi \in C_c^1(\Omega, \mathbb{R}^2)^n$, where C_c^1 denotes the set of smooth functions with compact support. Then the perimeter of Ω_i is given as

$$\begin{aligned} \text{Per}_g(\Omega_i, \Omega) &= TV_g(u_i) = \int_{\Omega} g |Du_i| \\ &= \sup_{\xi_i: |\xi_i| \leq g} \left\{ - \int_{\Omega} u_i \operatorname{div} \xi_i \, dx \right\}. \end{aligned} \quad (3)$$

Here Du_i refers to the distributional derivative of the function u_i , which is also defined for non-smooth functions. The representation on the right arises through integration by parts and relies on the dual variables ξ_i . It is valid for any L_1 -integrable function (including characteristic functions of sets with measurable boundary).

Minimization of total variation based geometric optimization problems is typically done by convex relaxation techniques. To this end, one ignores the binary constraint on the indicator variables u_i , optimizes respective energies over the convex function space $BV(\Omega, [0, 1])$ and subsequently binarizes the computed real-valued solutions. Such total-variation based approaches to partitioning problems were proposed in Chan et al. (2006), Cremers et al. (2004) and Cremers et al. (2006).

4.3 Convex Relaxations of the Multi-label Segmentation Problem

For the multi-label segmentation problem three convex relaxations of the minimal partition problem (1) were proposed in the fall of 2008 by Lellmann et al. (2009), by Zach et al. (2008) and by Chambolle et al. (2008). These approaches differ in the norm used in the regularizer, in their computational complexity and in the tightness of the relaxation. Since the latter two relaxations were shown to be tighter (Chambolle et al. 2008), we will concentrate on these in the following.

Zach et al. (2008) and Chambolle et al. (2008) represent the n regions Ω_i by the indicator function u in (2) which transforms the minimal partition problem (1) into the equivalent form

$$\begin{aligned} \min_{u \in \text{BV}(\Omega, \{0,1\})^n} & \left\{ \sum_{i=1}^n \int_{\Omega} u_i f_i \, dx + \frac{\lambda}{2} \int_{\Omega} g |Du_i| \right\} \\ \text{s.t.} & \sum_{i=1}^n u_i(x) = 1, \quad \forall x \in \Omega. \end{aligned} \quad (4)$$

As shown in (3), the total variation can be rewritten based on the dual variables ξ_i . Relaxing the binary variables u_i to real-valued ones, we obtain the relaxed convex problem

$$\begin{aligned} \min_{u \in \mathcal{B}} \sup_{\xi \in \mathcal{K}^{\lambda g}} & \left\{ \sum_{i=1}^n \int_{\Omega} u_i f_i \, dx - \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div} \xi_i \, dx \right\} \\ \mathcal{B} &= \left\{ u \in \text{BV}(\Omega, [0, 1])^n \mid \sum_{i=1}^n u_i(x) = 1, \quad \forall x \in \Omega \right\}. \end{aligned} \quad (5)$$

The difference between the relaxations by Zach et al. (2008) and Chambolle et al. (2008) only lies in the definition of the constraint set $\mathcal{K}^{\lambda g}$.

4.3.1 ZGFN-Relaxation

The relaxation by Zach et al. (2008) is based on the energy in (5) with the following constraints on the dual variables according to the dual formulation of the total variation in (3)

$$\begin{aligned} \mathcal{K}_{\text{ZGFN}}^{\lambda g} &= \left\{ \xi \in C_c^1(\Omega, \mathbb{R}^2)^n \mid |\xi_i(x)| \leq \frac{\lambda g(x)}{2}, \right. \\ & \left. 1 \leq i \leq n, \quad \forall x \in \Omega \right\}. \end{aligned} \quad (6)$$

4.3.2 CCP-Relaxation

Chambolle et al. (2008) originally proposed an energy formulation based on the upper level sets, which can be transformed into the energy formulation in (5) with the following constraint set on the dual variables where $\xi_{n+1} = 0$

$$\begin{aligned} \mathcal{K}_{\text{CCP}}^{\lambda g} &= \left\{ \xi \in C_c^1(\Omega, \mathbb{R}^2)^{n+1} \mid |\xi_i(x) - \xi_j(x)| \leq \lambda g(x), \right. \\ & \left. 1 \leq i < j \leq n+1, \quad \forall x \in \Omega \right\}. \end{aligned} \quad (7)$$

4.4 Optimization

For the indicated continuous relaxations we give update schemes, which are iterated until convergence. They are based on the primal–dual algorithm proposed by Pock et al. (2009) with step sizes chosen according (Pock and Chambolle 2011).

4.4.1 ZGFN-Relaxation

Optimization of Sect. 4.3.1 is done by iterating the following update scheme. Essentially, it consists of alternating a projected gradient descent in the primal variables u with projected gradient ascent in the dual variables ξ . In addition, it contains an over-relaxation step in the primal variables (giving rise to auxiliary variables \bar{u}) assuring fast convergence of the algorithm (Pock et al. 2009):

$$\begin{aligned} \xi^{t+1} &= \Pi_{\mathcal{K}_{\text{ZGFN}}^{\lambda g}} \left(\xi^t + \frac{1}{2} \nabla \bar{u}^t \right) \\ u^{t+1} &= \Pi_{\mathcal{B}} \left(u^t - \frac{1}{4} (f - \operatorname{div} \xi^{t+1}) \right) \\ \bar{u}^{t+1} &= u^{t+1} + (u^{t+1} - u^t) = 2u^{t+1} - u^t. \end{aligned} \quad (8)$$

The projection $\Pi_{\mathcal{K}_{\text{ZGFN}}^{\lambda g}}$ of the dual variables ξ is carried out by simple clipping, while that of the primal variables $\Pi_{\mathcal{B}}$ is a projection onto the simplex in \mathbb{R}^n (Michelot 1986).

4.4.2 CCP-Relaxation

In order to find the globally optimal solution to the relaxed convex optimization problem in Sect. 4.3.2 we introduce the constraints on the dual variables (7) into the energy in (5) by means of Lagrange multipliers $\mu_{ij} : \Omega \rightarrow \mathbb{R}^2$ and additional variables $q_{ij} := \xi_i - \xi_j \in \mathcal{Q}$

$$\mathcal{Q} := \left\{ q_{ij} \in C_c^1(\Omega, \mathbb{R}^2) \mid |q_{ij}(x)| \leq \lambda g(x), \right. \\ \left. 1 \leq i < j \leq n + 1, \forall x \in \Omega \right\}. \tag{9}$$

We obtain

$$\min_{\substack{u \in \mathcal{B} \\ \mu_{ij} \in \mathcal{M} \\ q_{ij} \in \mathcal{Q}}} \sup_{\substack{\xi_i \in C_c^1 \\ q_{ij} \in \mathcal{Q}}} \left\{ \sum_{i=1}^n \int_{\Omega} u_i (f_i - \operatorname{div} \xi_i) \, dx \right. \\ \left. + \sum_{1 \leq i < j \leq n+1} \int_{\Omega} \mu_{ij} (\xi_i - \xi_j - q_{ij}) \, dx \right\}. \tag{10}$$

Note again that $\xi_{n+1} = 0$. As before we employ the primal–dual algorithm published in Pock et al. (2009) based on the following update scheme:

$$\xi^{t+1} = \xi^t + \frac{1}{2+n} \left(\nabla \bar{u}^t + \left(\sum_{j>i} \bar{\mu}_{ij}^t - \sum_{i>j} \bar{\mu}_{ji}^t \right)_{1 \leq i \leq n} \right) \\ q^{t+1} = \Pi_{\mathcal{Q}} \left(q^t + (-\bar{\mu}) \right) \\ u^{t+1} = \Pi_{\mathcal{B}} \left(u^t - \frac{1}{4} (f - \operatorname{div} \xi^{t+1}) \right) \\ \mu^{t+1} = \mu^t - \frac{1}{3} (\xi_i - \xi_j - q_{ij}) \\ \bar{u}^{t+1} = 2u^{t+1} - u^t \\ \bar{\mu}^{t+1} = 2\mu^{t+1} - \mu^t. \tag{11}$$

The projections of the dual variables $\Pi_{\mathcal{Q}}$ are carried out by simple clipping, while that of the primal variables $\Pi_{\mathcal{B}}$ is a projection onto the simplex in \mathbb{R}^n (Michelot 1986).

4.5 Implementation and Optimality

For the implementation of the continuous approaches, information on the discretization, the convergence, termination criteria, the computation of the binary solutions based on the relaxed solution and the pseudocode of the algorithms is given in the following.

4.5.1 Discretization

For optimization of the given relaxations, the energies have to be discretized. The computation of the gradient of the primal variables, ∇u_i , is done by forward differences with von Neumann boundary conditions. Let w denote the width

of the image and h its height. Then the gradient is given by

$$\nabla u_i(x, y) = \begin{cases} (u_i(x+1, y) - u_i(x, y), & \text{if } x < w, y < h \\ u_i(x, y+1) - u_i(x, y)) & \text{if } x = w, y < h \\ (0, u_i(x, y+1) - u_i(x, y)) & \text{if } x < w, y = h \\ (u_i(x+1, y) - u_i(x, y), 0) & \text{if } x = w, y = h \\ (0, 0) & \end{cases}$$

The divergence of the dual variables ξ_i is computed by backward differences

$$\operatorname{div} \xi_i(x, y) = \begin{cases} \xi_i^1(x, y) - \xi_i^1(x - 1, y) & \text{if } 1 < x < w \\ \xi_i^1(x, y) & \text{if } x = 1 \\ -\xi_i^1(x - 1, y) & \text{if } x = w \\ \xi_i^2(x, y) - \xi_i^2(x, y - 1) & \text{if } 1 < y < h \\ \xi_i^2(x, y) & \text{if } y = 1 \\ -\xi_i^2(x, y - 1) & \text{if } y = h \end{cases}$$

4.5.2 Convergence and Termination Criteria

There are different ways to determine the convergence of the algorithm. Ideally, the primal–dual gap can be computed, which is the difference between the primal and the dual energy of the optimization problem. This is not always possible as the primal and dual energies themselves are optimization problems of variable complexity. Exemplarily, we give the primal energy of the optimization problem by (4):

$$E_p(u) = \left\{ \sum_{i=1}^n \int_{\Omega} u_i f_i \, dx - \lambda \int_{\Omega} g |D u_i| \right\} + \delta_{\mathcal{B}}(u), \tag{12}$$

where $\delta_{\mathcal{B}}$ is the indicator function of the set \mathcal{B} : $\delta_{\mathcal{B}}(u) = \begin{cases} 0, & u \in \mathcal{B} \\ \infty, & u \notin \mathcal{B} \end{cases}$. Computation of the dual energy amounts to a point-wise optimization problem

$$E_d(\xi) = \min_{u \in \mathcal{B}} \left\{ \sum_{i=1}^n \int_{\Omega} u_i (f_i - \operatorname{div} \xi_i) \, dx - \delta_{\mathcal{K}_{\text{ZGFN}}^{\lambda g}}(\xi) \right\} \\ = \int_{\Omega} \min_i (f_i - \operatorname{div} \xi_i) \, dx - \delta_{\mathcal{K}_{\text{ZGFN}}^{\lambda g}}(\xi). \tag{13}$$

Here $\delta_{\mathcal{K}_{\text{ZGFN}}^{\lambda g}}$ denotes the indicator function for the set $\mathcal{K}_{\text{ZGFN}}^{\lambda g}$. During the optimization, the primal energy decreases, while the dual energy increases. When the optimal solution is reached, the primal–dual gap goes to zero. The size of the gap can be used to formulate suitable convergence criteria for the algorithm, e.g. if the gap decreased less than 1% from one iteration to the next. Figure 1 shows the primal–dual gap for the ZGFN algorithm for up to 2,500 iterations.

As the primal–dual gap is difficult to compute for the CCP-relaxation, we apply a different convergence criterion in this survey. It is based on the difference between the current and the previous relaxed solution of the optimization problem.

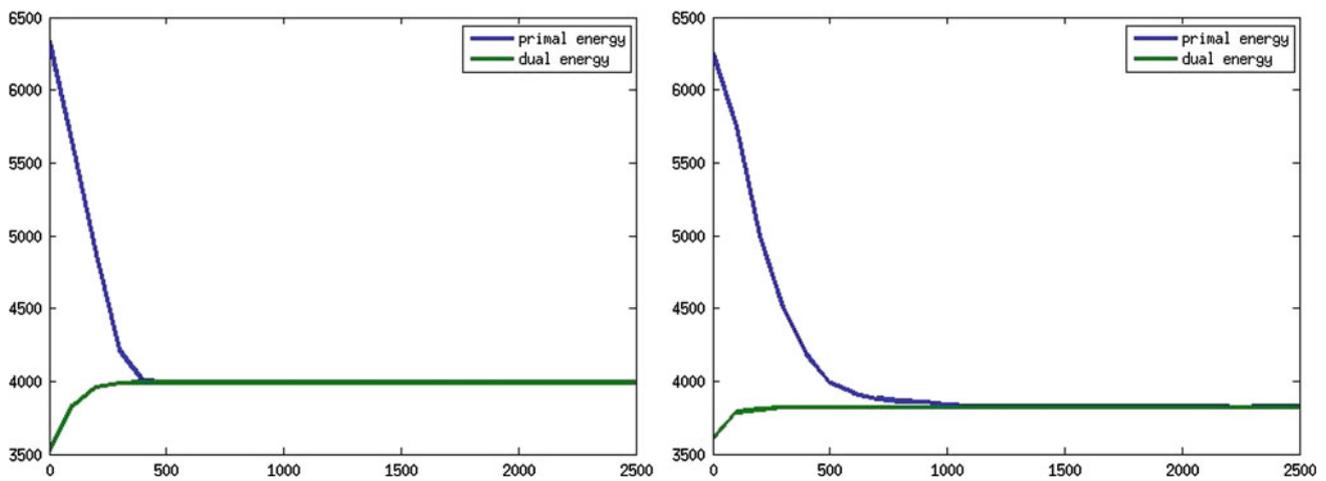


Fig. 1 Convergence analysis of the algorithm by Zach et al. for two examples from the Graz benchmark database. The progress of the primal energy (blue) is depicted with respect to the dual energy (green).

If the norm of this difference vector averaged over the number of pixels falls below a certain threshold we regard the optimization procedure as converged. In this paper we use 0.0001 as the threshold.

4.5.3 Binarization and Optimality Bounds

To obtain a labeling function $l : \Omega \rightarrow \{1, \dots, n\}$ and a binary indicator function $u : \Omega \rightarrow \{0, 1\}^n$ from the globally optimal relaxed solution $\tilde{u} \in \mathcal{B}$ some rounding scheme must be applied, which can lead to suboptimal results. We assign each pixel x to the first label with maximum indicator function \tilde{u}_i

$$l(x) = \min_i \{\arg \max \{\tilde{u}_i(x)\}\},$$

$$u_i(x) = \begin{cases} 1, & l(x) = i \\ 0, & \text{otherwise} \end{cases} \quad \forall x \in \Omega. \quad (14)$$

There are other rounding schemes, which can be more appropriate e.g. in the case of anisotropic total variation regularization (Lellmann et al. 2009).

Chan et al. (2006) proved the thresholding theorem for total variation approaches, which states that for the case of two segments (i.e. $n = 2$ with a single indicator function $u : \Omega \rightarrow \mathbb{R}$) any threshold $T \in (0, 1)$ can be applied to obtain a globally optimal binary solution. No such theorem can be proved for the case of more than two segments. Yet, in the multiregion case we can compute bounds of optimality.

Let $u^* \in \mathcal{B}$ be the global minimizer of the original binary problem formulation (4), $\tilde{u} \in \mathcal{B}$ the solution of the relaxed problem (5) and u the binarized solution (14) of the relaxed problem. Chambolle et al. (2008) gave an a posteriori optimality bound, which states that

The optimal solution of the relaxed optimization problem is obtained as soon as the primal–dual gap goes to zero (Color figure online)

$$E(u) - E(u^*) \leq E(u) - E(\tilde{u}). \quad (15)$$

This means that the energy difference between the globally optimal binary result and the binarized solution of the algorithm is bounded by the energy difference between the relaxed and the binarized solution.

In (Lellmann et al. 2011), an a priori optimality bound based on a probabilistic rounding scheme was given stating that

$$\mathbb{E}(E(u)) \leq 2E(u^*) \quad (16)$$

for the Potts model. A similar “factor of two”-bound was earlier suggested in the discrete setting by Boykov et al. (2001) (see Sect. 5.3.3).

4.5.4 Pseudocode

The pseudocode for the algorithm by Pock et al. (2009) is given as Algorithm 1. The only input parameters are the data term f , the smoothness parameter λ and the stopping parameter ϵ .

Algorithm 1 Continuous Multi-label-Segmentation

```

1: procedure OPTIMIZE( $f, \lambda, \epsilon$ )
2:   Initialize  $u^0 \in [0, 1]^n$  arbitrarily,  $t := 1$ 
3:   repeat
4:      $t = t + 1$ 
5:     Compute primal–dual updates as in (8) or (11)
6:     Compute difference vector  $d = u^t - u^{t-1}$  (Sect. 4.5.2)
7:   until  $|d| < \epsilon|\Omega|$  ▷ Stopping criterion
8:   Compute labeling  $l$  as in (14)
9:   return labeling
10: end procedure

```

4.6 Related Approaches

There have been several extensions to the presented TV-based segmentation approaches. Examples include weighted TV norms which depend on the label indices (Lellmann et al. 2009), anisotropic TV norms (Kolev et al. 2010) and multi-label approaches which allow direction-dependent transition penalties (Stekalovskiy and Cremers 2011).

5 The Spatially Discrete Setting

In this section we discuss discrete multi-label segmentation based on the minimization of the Potts energy derived from an MRF model. We review different relaxations as well as minimization and inference algorithms and concentrate on the best performing methods today, α -expansion and Fast-PD.

5.1 Segmentation via Inference on Markov Random Fields

In the discrete setting the segmentation task can be formulated as an undirected graph, a so called Markov random field. The nodes of the graph represent random variables taking on different labels whereas the edges encode the dependencies between different variables.

The multi-label segmentation problem can then be formulated as the minimization of the following energy, which is a discrete version of (1)

$$\min_u \sum_{x \in \Omega} \sum_{i \in \mathcal{L}} f_{ix} u_{ix} + \frac{\lambda}{2} \sum_{x \sim y} g_{xy} \sum_{i \in \mathcal{L}} |u_{ix} - u_{iy}| \quad (17)$$

$$s.t. \sum_{i \in \mathcal{L}} u_{ix} = 1, \quad \{u_{ix}\} \in \{0, 1\}^{|\mathcal{L} \times \Omega|}.$$

Here, $x \sim y$ denotes that x and y are neighboring nodes in the grid. Typically four or eight pixel neighborhoods are considered.

The unary terms determine how well each pixel complies with each label, where f_{ix} indicates the cost for assigning pixel $x \in \Omega$ to label $i \in \mathcal{L}$. The pairwise terms impose some sort of regularity on neighboring labels. They are weighted by the edge indicator function $g_{xy} = \frac{g(x)+g(y)}{2}$ (see explanation following (1)).

We will present two linear relaxation schemes for the above energy formulation, which are directly related to the continuous ZGFN- and CCP-relaxations.

5.1.1 KT-Relaxation

The following linear relaxation of the energy in (17) was proposed by Kleinberg and Tardos (2002):

$$\min_u \sum_{x \in \Omega} \sum_{i \in \mathcal{L}} f_{ix} u_{ix} + \frac{\lambda}{2} \sum_{x \sim y} g_{xy} \sum_{i \in \mathcal{L}} u_{ixy} \quad (18)$$

$$s.t. \sum_{i \in \mathcal{L}} u_{ix} = 1, \quad u_{ixy} \geq |u_{ix} - u_{iy}|, \quad u_{ix} \geq 0, \quad u_{ixy} \geq 0$$

The indicator variables in their binarized form, $u_{ix} \in \{0, 1\}^{|\mathcal{L} \times \Omega|}$ and $u_{ixy} \in \{0, 1\}^{|\mathcal{L} \times \Omega \times \Omega|}$, can be understood as follows:

$$u_{ix} = \begin{cases} 1, & l(x) = i \\ 0, & \text{otherwise} \end{cases} \quad u_{ixy} = \begin{cases} 1, & l(x) = i \oplus l(y) = i \\ 0, & \text{otherwise,} \end{cases}$$

where \oplus means exclusive or.

5.1.2 S-Relaxation

A more general linear relaxation of the energy in (17) was suggested by Schlesinger (1976) and later revisited by Werner (2007):

$$\min_u \sum_{x \in \Omega} \sum_{i \in \mathcal{L}} f_{ix} u_{ix} + \frac{\lambda}{2} \sum_{x \sim y} g_{xy} \sum_{i, j \in \mathcal{L}} u_{ijxy} \quad (19)$$

$$s.t. \sum_{i \in \mathcal{L}} u_{ix} = 1, \quad u_{ix} = \sum_{j \in \mathcal{L}} u_{ijxy} \quad \forall y,$$

$$u_{jy} = \sum_{i \in \mathcal{L}} u_{ijxy} \quad \forall x, \quad u_{ix} \geq 0, \quad u_{ijxy} \geq 0.$$

The indicator variables in their binarized form, $u_{ix} \in \{0, 1\}^{|\mathcal{L} \times \Omega|}$ and $u_{ijxy} \in \{0, 1\}^{|\mathcal{L} \times \mathcal{L} \times \Omega \times \Omega|}$, can be understood in a slightly different way

$$u_{ix} = \begin{cases} 1, & l(x) = i \\ 0, & \text{otherwise} \end{cases} \quad u_{ijxy} = \begin{cases} 1, & l(x) = i \wedge l(y) = j \\ 0, & \text{otherwise.} \end{cases}$$

5.2 Optimization

For the Potts model with two labels a polynomial-time optimal solution of the MRF problem can be computed by solving a minimum s-t-cut (Greig et al. 1989). For more than two labels, the problem is NP-hard (Boykov et al. 2001). Therefore, efficient algorithms only allow for approximate solutions with corresponding optimality bounds with respect to the global optimum. Several inference algorithms exist to compute such approximations:

5.2.1 Message passing

Describes iterative algorithms for computing MAP estimates or marginals in Markov Random Fields. For calculating MAP estimates the *max-product* (or *min-sum*) algorithm can be applied. For computing marginals, the *sum-product*

algorithm is used of which *Belief Propagation* (Pearl 1988) is a special case. In general, message passing strategies find the global optimum only for tree shaped graphs. Newer algorithms for graphs containing cycles were proposed by Wainwright et al. (2005) and Kolmogorov (2006). A message passing scheme based on *dual decomposition* was proposed by Komodakis et al. (2007). It decomposes the original problem into a set of simpler optimization problems under the assumption that their solutions are consistent. If the domain is decomposed into trees the algorithm is parallelizable. However, the smaller the trees the slower is the convergence of the algorithm. Hence, qualitatively this parallelization is not comparable to that obtained by the continuous algorithms on the GPU.

Apart from message passing, algorithms that iteratively change one or several node labels in order to improve the solution are often applied for optimization. *Iterated conditional modes* (Besag 1986) (ICM) and *simulated annealing* (Geman and Geman 1984) only allow for label changes of single nodes in each iteration leading to local minima in the former case and an exponential runtime in the latter. Algorithms that change labels of multiple nodes include α - β -swap, α -expansion and FastPD, all of which are optimized by iteratively solving a set of graph cut problems.

Szeliski et al. (2006) compared different algorithms for the minimization of Markov random field approaches: ICM, graph cuts (α -expansion and α - β -swap) and belief propagation algorithms. Of those algorithms, α -expansion yields the best performance results for the Potts model. Hence, in this paper we concentrate on α -expansion and its recent generalization FastPD.

Both algorithms approximate the Potts model segmentation problem (17) based on graph cuts. They do not compute a relaxation of the minimization problem in the sense of Kleinberg and Tardos or Schlesinger, but instead the original functional (17) is optimized under relaxed optimality conditions, see Sect. 5.2.3.

In computer vision, graph cuts were applied for the first time by Greig et al. (1989). The basic idea is to compute a binary partition ('cut') of the nodes in a graph that separates a *source* from a *sink* node. The sum of the edge weights which are cut by the partition interface should be minimal over all possible separations. This is equivalent to computing a maximum flow on the graph, which can be done in polynomial time.

A globally optimal graph cut solution for multi-label segmentation problems was proposed by Ishikawa (2003) for the case of a convex regularizer and a linear label space. A related convex variational approach was proposed by Pock et al. (2010). Unfortunately the Ishikawa approach does not apply to discontinuity preserving energies such as the Potts model.

5.2.2 α -Expansion

In 2001, Boykov et al. (2001) proposed to solve multi-label segmentation problems by repeated graph cuts. Each step of the algorithm iterates over all labels α . For each α an 'expansion move' is computed in which each pixel can either switch to α or keep its current label. The procedure terminates if no expansion moves exist that reduce the energy.

Two advantages come with α -expansion: firstly, for each label α the *optimal* expansion move can be computed given that the pairwise term is a metric in the label space. Secondly, the solution lies energetically within a constant factor from the global optimum (see Sect. 5.3.3). Other strategies for label changes, such as α - β -swap, have no provable optimality bounds and do not perform as well in practice.

5.2.3 Fast Primal–Dual

A generalization of α -expansion was proposed by Komodakis and Tziritas (2005). They suggested a primal–dual technique to minimize the energy in (17) based on the S-relaxation (19). The relaxed linear program and its dual are of the following form with suitable A , b and c for (19)

$$\begin{aligned} \text{primal: } \min c^T u, \quad & \text{s.t. } Au = b, \quad x \geq 0 \\ \text{dual: } \max b^T v, \quad & \text{s.t. } A^T v \leq c. \end{aligned} \quad (20)$$

In contrast to other LP-relaxation techniques, Komodakis et al. do not solve the relaxed linear program directly, but instead generate a sequence of integer-primal and relaxed-dual solutions in each step. The basic idea is that if the *relaxed dual complementary slackness conditions*

$$\forall u_i > 0 \implies \frac{c_i}{h} \leq \sum_{j=1}^m a_{ji} v_j \leq c_i \quad (21)$$

are fulfilled, then the primal (integer) solution u is an approximation of the global (integer) optimum u^* with $c^T u \leq h \cdot c^T u^*$ where $h := \max_i h_i$.¹ For the Potts model h equals two.

Algorithmically, FastPD works similarly to α -expansion: each step iterates over all labels and for each label computes a new pair of integer-primal, relaxed-dual solutions by solving a graph cut problem. If no more label changes occur the algorithm terminates and the solution is shown to fulfill (20).

FastPD and α -expansion are equivalent in the sense that both algorithms compute the optimal α -expansion in each step leading to identical results and optimality bounds. However, in contrast to α -expansion, the Fast-PD method keeps

¹ In Komodakis and Tziritas (2005) several algorithms are proposed for different choices of parameters h_i . In this paper we use the α -expansion equivalence of FastPD (called PD $_{2,\mu=1}$ by Komodakis and Tziritas) since it corresponds to the Potts model.

track of the relaxed dual energy which leads to a simplified graph cut problem in each step and decreases the overall runtime. In the following, the FastPD-method will be called FastPD4 and FastPD8 depending on the size of the selected pixel neighborhood.

5.3 Implementation and Optimality

In the following we will give details on the implementation of the discrete approaches and discuss optimality bounds for solutions.

5.3.1 Label Selection Strategy

Different orders are conceivable for the selection of labels for α -expansion and FastPD. We will show in the experimental Sect. 7 that the segmentation results as well as the runtime are influenced by this order. α -expansion maintains a priority list of labels for which a previous expansion step led to an energy reduction. In contrast, FastPD uses a fixed label order in each cycle.

5.3.2 Pseudocode

For testing the discrete methods we use the following publicly available implementations: the α -expansion code by Veksler and Delong,² which uses the graph cut implementation by Boykov and Kolmogorov (2004), and the FastPD code implemented by Komodakis and Tziritas (2005).³ Pseudocode for both algorithms are given in Algorithms 2 and 3.

Algorithm 2 α -Expansion

```

1: procedure OPTIMIZE( $f, \lambda$ )
2:   Initialize labeling  $\mathbf{u}$  arbitrarily
3:   labelChanged = false
4:   for all  $\alpha \in \mathcal{L}$  do            $\triangleright$  Label priority list, see Sect. 5.3.1
5:     Compute the  $\alpha$ -expansion  $\hat{\mathbf{u}}$  with lowest energy  $E(\hat{\mathbf{u}})$  via
       graph cuts
6:     if  $E(\hat{\mathbf{u}}) < E(\mathbf{u})$  then
7:        $\mathbf{u} = \hat{\mathbf{u}}$ 
8:       labelChanged = true
9:     end if
10:  end for
11:  if labelChanged then
12:    goto 3
13:  end if
14:  return labeling
15: end procedure

```

² <http://vision.csd.uwo.ca/code/>

³ <http://www.csd.uoc.gr/~komod/FastPD/>

Algorithm 3 Fast Primal Dual Algorithm

```

1: procedure OPTIMIZE( $f, \lambda$ )
2:   Initialize labeling  $\mathbf{u}$  arbitrarily
3:   Initialize dual variables  $\mathbf{v}$  according to the dual constraints
4:   for all  $\alpha \in \mathcal{L}$  do            $\triangleright$  Linear label list, see Sect. 5.3.1
5:     Pre-edit dual variables  $\mathbf{v}$  (Komodakis et al. 2007)
6:     Compute new integer-primal and dual pair  $(\hat{\mathbf{u}}, \hat{\mathbf{u}})$  via graph
       cuts
7:     Post-edit dual variables  $\hat{\mathbf{u}}$  (Komodakis et al. 2007)
8:     if  $\hat{\mathbf{u}} \neq \mathbf{u}$  then
9:        $(\mathbf{u}, \mathbf{v}) = (\hat{\mathbf{u}}, \hat{\mathbf{u}})$ 
10:      labelChanged = true
11:    end if
12:  end for
13:  if labelChanged then
14:    goto 4
15:  end if
16:  return labeling
17: end procedure

```

5.3.3 Optimality Bounds

Both α -expansion and Fast-PD can be shown to yield solutions lying within specific a priori optimality bounds, e.g. for the Potts model

$$E(l) \leq 2E(l^*), \quad (22)$$

Here l is the computed solution and l^* is a globally optimal solution of (17). Note that this bound is the discrete equivalent to the bound shown by Lellmann et al. in (16) for the continuous multi-label minimal partition problem.

Similar to the continuous methods, an a posteriori bound can also be computed for the discrete approaches. Let \tilde{l} be the solution of the relaxed problem, then we can state that

$$E(l) - E(l^*) \leq E(l) - E(\tilde{l}). \quad (23)$$

5.4 Related Approaches

There are several extensions to the presented approaches, e.g. the LogCut algorithm (Lempitsky et al. 2007) which fuses the current solution with an alternate solution in each step, ordering constraints enforcing spatial relationships between labels (Liu et al. 2010; Felzenszwalb and Veksler 2010), anisotropic regularization based on Wulff-shapes (Zach et al. 2009) and performance improvements (Alahari et al. 2010; Batra and Kohli 2011). Veksler (2007, 2009) proposed *multi-label moves* to minimize multi-label segmentation energies with truncated convex priors resulting in piecewise smooth solutions.

6 A Theoretical Comparison of Continuous and Discrete Multi-label Approaches

We have presented several relaxations of the Potts model suggested in the MRF and PDE communities. We will now characterize their relation for L_1 regularity ($\int_{\Omega} |\nabla u|_1 \, dx$) and L_2 regularity ($\int_{\Omega} |\nabla u|_2 \, dx$).

Proposition 1 *For L_1 regularity the convex relaxations by Zach et al. in 4.3.1, Chambolle et al. in 4.3.2, Kleinberg and Tardos in 5.1.1 and Schlesinger in 5.1.2 are equivalent.*

Proof 1. Equivalence of the relaxations by Zach et al. and Chambolle et al. for L_1 regularity

For one-dimensional problems this equivalence was shown by Chambolle et al. (2008). We generalize this proof to L_1 regularity by showing that the minimum energies for both relaxations are equal. Remember that both relaxations merely differ in the constraint set of the dual variables, i.e. the sets $\mathcal{K}_{ZGFN}^{\lambda g}$ in (6) and $\mathcal{K}_{CCP}^{\lambda g}$ in (7). For simplicity, we will omit the dependency on x .

Let $\xi \in \mathcal{K}_{ZGFN}^{\lambda g}$. The dual norm of the L_1 norm is the maximum norm, so for all elements $1 \leq i \leq n$ we have $|\xi_i|_{\infty} \leq \frac{\lambda g}{2}$. For any two indices i and j it follows that $|\xi_i - \xi_j|_{\infty} \leq \lambda g$ and hence $\xi \in \mathcal{K}_{CCP}^{\lambda g}$. This implies that $\mathcal{K}_{ZGFN}^{\lambda g} \subset \mathcal{K}_{CCP}^{\lambda g}$ and thus $E_{ZGFN}(u, \xi) \geq E_{CCP}(u, \xi)$.

Let now $\hat{\xi} \in \mathcal{K}_{CCP}^{\lambda g}$, i.e. $|\hat{\xi}_i - \hat{\xi}_j|_{\infty} \leq \lambda g$. We will now define a translation which maps each $\hat{\xi} \in \mathcal{K}_{CCP}^{\lambda g}$ to an element $\xi \in \mathcal{K}_{ZGFN}^{\lambda g}$ without affecting the energy—see below. To this end, we define $\xi \in C_c^1(\Omega, \mathbb{R}^2)^n$

$$\xi_i := \hat{\xi}_i - C, \quad C := \frac{\max_i \hat{\xi}_i + \min_i \hat{\xi}_i}{2} \in \mathbb{R}^2, \tag{24}$$

where min/max denote the component-wise minimum/maximum. We obtain

$$\begin{aligned} |\xi_i|_{\infty} &= \max \left\{ |\hat{\xi}_i^1|, |\hat{\xi}_i^2| \right\} \\ &= \max \left\{ \left| \hat{\xi}_i^1 - C_1 \right|, \left| \hat{\xi}_i^2 - C_2 \right| \right\} \leq \frac{\lambda g}{2}. \end{aligned} \tag{25}$$

It follows that $\xi \in \mathcal{K}_{ZGFN}^{\lambda g}$.

The constant shift introduced in (23) does not affect the energy. To see this, it suffices to consider the energy term measuring the contour length, since the data term is independent of ξ . Due to the Divergence Theorem of Gauss we have

$$\begin{aligned} E_{TV}(u, \xi) &= \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div} \xi_i \, dx \\ &= E_{TV}(u, \hat{\xi}) - \int_{\Omega} \sum_{i=1}^n u_i \operatorname{div} C \, dx = E_{TV}(u, \hat{\xi}) \end{aligned} \tag{26}$$

since $\sum_i u_i = 1$ and $\hat{\xi} \in C_c^1(\Omega, \mathbb{R}^2)^n$. It follows that $E_{CCP}(u, \hat{\xi}) \geq E_{ZGFN}(u, \xi) \forall \hat{\xi} \in \mathcal{K}_{CCP}$. Altogether, we obtain $\min_{u, \xi} E_{CCP}(u, \xi) = \min_{u, \xi} E_{ZGFN}(u, \xi)$. This proves the equivalence of both relaxations in the case of L_1 regularity.

2. Equivalence of the relaxations by Zach et al. and Kleinberg and Tardos for L_1 regularity

We now show the equivalence between the relaxed optimization problems in (18) and (4) (with relaxed $u : \Omega \rightarrow [0, 1]^n$). Kleinberg and Tardos minimize over the variables u_{ixy} , which only appear in the second energy term defining the contour length. Thus, in the minimum the constraint $u_{ixy} \geq |u_{ix} - u_{iy}|$ holds with equality and we obtain the following energy

$$\begin{aligned} \min_u \sum_{x \in \Omega} \sum_{i \in \mathcal{L}} f_{ix} u_{ix} + \frac{\lambda}{2} \sum_{x \sim y} g_{xy} \sum_{i \in \mathcal{L}} |u_{ix} - u_{iy}| \\ \text{s.t.} \quad \sum_{i \in \mathcal{L}} u_{ix} = 1, \quad u_{ix} \geq 0. \end{aligned} \tag{27}$$

For a 4-neighborhood relation for $x \sim y$, this energy is equivalent to the discretization of the energy in (4) after relaxation of the indicator function u .

3. Equivalence of the relaxations by Kleinberg and Tardos and by Schlesinger

This equivalence was shown by Osokin et al. (2011).

Interestingly, the above equivalence of various relaxations is no longer true for L_2 regularity. \square

Proposition 2 *For L_2 regularity the linear relaxations of Kleinberg and Tardos and Schlesinger no longer apply. Moreover, the relaxation by Chambolle et al. is tighter than that by Zach et al.*

Proof In the case of L_2 regularity the linear relaxations by Kleinberg and Tardos and by Schlesinger are not applicable since they do not admit quadratic constraints. Among the convex relaxations the one by Chambolle et al. dominates the one by Zach et al.—see also Chambolle et al. (2008): W.l.o.g. we assume $\lambda g = 1$. While we still have $\mathcal{K}_{ZGFN}^1 \subset \mathcal{K}_{CCP}^1$, the converse no longer holds. We indicate an element $\xi \in \mathcal{K}_{CCP}^1$ which does not admit a corresponding element $\hat{\xi} \in \mathcal{K}_{ZGFN}^1$:

Let $\xi = (\xi_1, \xi_2, \xi_3, 0, \dots, 0) \in \mathcal{K}_{CCP}^1$ with $\xi_1 = (-\frac{1}{2}, 0)$, $\xi_2 = (\frac{1}{2}, 0)$, $\xi_3 = (0, \frac{\sqrt{3}}{2})$. As shown in Fig. 2, these vectors form an isosceles triangle with edge length one. A corresponding $\hat{\xi} \in \mathcal{K}_{ZGFN}^1$ must preserve these distances and, thus, must fulfill:

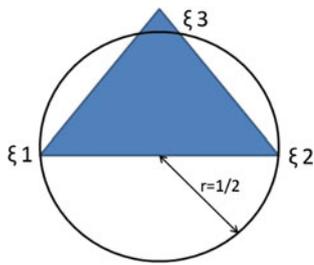


Fig. 2 For L_2 regularity, the CCP-relaxation dominates the ZGFN-relaxation because the set of permissible dual variables is strictly larger: $\mathcal{K}_{ZGFN}^1 \subsetneq \mathcal{K}_{CCP}^1$. The three corners of the blue isosceles triangle define an element $\xi = (\xi_1, \xi_2, \xi_3, 0, \dots, 0) \in \mathcal{K}_{CCP}^1$ which cannot be mapped to a corresponding element $\hat{\xi} \in \mathcal{K}_{ZGFN}^1$, because there exists no circle of radius $1/2$ which contains this triangle

$$|\hat{\xi}_i| \leq \frac{1}{2} \quad \text{and} \quad |\hat{\xi}_1 - \hat{\xi}_2| = |\hat{\xi}_1 - \hat{\xi}_3| = |\hat{\xi}_2 - \hat{\xi}_3| = 1. \tag{28}$$

Clearly, there exists no circle with radius $\frac{1}{2}$ which contains all three vertices $\hat{\xi}_i$ of this isosceles triangle. \square

From this follows that the CCP-relaxation dominates the ZGFN-relaxation for L_2 regularity, i.e. $E_{CCP} \geq E_{ZGFN}$, and thus the solution of the relaxed problem is energetically closer or equally close to the global optimum of the binary problem. However, the ZGFN-relaxation is simpler than the CCP-relaxation since it leads to simpler constraints on the dual variables. In addition, it can be implemented more efficiently because the number of constraints on the dual variables (6) increases only linearly with the number of labels. In contrast, the CCP-relaxation imposes quadratically many constraints on the dual variables (7) yielding higher computation times. In practice, one can accelerate the CCP-relaxation by first solving the ZGFN-relaxation and locally switching to the CCP-relaxation if needed, i.e. if the boundary of the smaller constraint set is reached (Zach et al. 2012).

7 An Experimental Comparison of Discrete and Continuous Multi-label Approaches

We have introduced the main approaches for solving the minimal partition problem for multiple labels and have discussed relations, differences, advantages and drawbacks. To experimentally evaluate the performance of the different algorithms we apply them to the problem of interactive segmentation.

In this section we compare results of the proposed multi-label segmentation approaches for the relaxations by Zach et al. (PdeZGFN), by Chambolle et al. (PdeCCP), for α -expansion based on four and eight-neighborhood (Alpha4, Alpha8) as well as for the algorithm by Komodakis et al. (FastPD4, FastPD8). For the PDE approaches we set $\lambda = 67$. We used a GTX 680 for the parallel GPU computa-

tions. For α -expansion and Fast-PD we used $\lambda = 5.2$ for the 4-neighborhood case and $\lambda = 1.7$ for the 8-neighborhood case. Computations were carried out on an Intel Xeon Quad Core with 2.27GHz. All values for λ were chosen optimally for each approach with respect to the chosen data base. For segmentation we require a data term f , which indicates the association of each pixel with each label. Here, we use the data term for interactive segmentation proposed by Nieuwenhuis and Cremers (2012) with the parameters $\sigma = 1.2$, $\alpha = 0.7$ and $\gamma = 5$.

In the following, we will experimentally compare the different approaches with respect to segmentation quality, optimality and runtimes. Note that the energy formulations depend on model choices whose optima do not necessarily coincide with the ground truth of the segmentation. Hence the quality evaluation reflects both the quality of the optimization and of the chosen model.

7.1 Segmentation Accuracy

For unsupervised segmentation several benchmarks are available, e.g. the Berkeley database, the GrabCut database or the Pascal VOC Database. As extensively discussed by Santner (2010), these benchmarks are not suited for testing interactive segmentation. Hence, Santner et al. recently published the first benchmark for interactive scribble based multi-label segmentation containing 262 seed-groundtruth pairs from 158 natural images containing between 2 and 13 user labeled segments. The label frequencies (Table 1) are not uniformly distributed. Instead, small label numbers up to four appear frequently, whereas large numbers are rare. This influences the statistics on quality and runtime.

To assess the segmentation quality, the Dice-score relates the overlap area of the groundtruth $\bar{\Omega}_i$ and the computed segment Ω_i to their sum. Taking the arithmetic mean over all segments yields

$$Dice(\Omega, \bar{\Omega}) = \frac{1}{n} \sum_{i=1}^n \frac{2|\bar{\Omega}_i \cap \Omega_i|}{|\bar{\Omega}_i| + |\Omega_i|}. \tag{29}$$

The closer to one the Dice-score the more accurate is the segmentation. To evaluate the segmentation accuracy of the continuous and discrete approaches we computed the arithmetic mean and standard deviation of the Dice-score over all images of the database in Table 2. The values are very similar for all methods suggesting that the segmentation accuracy is independent of the optimization method.

7.1.1 Statistical Evaluation

We want to statistically validate this statement. To this end, we used the Kolmogorov-Smirnov test in order to test the following hypothesis for each two optimization methods

Table 1 Appearance frequency for different label numbers in the Graz benchmark for interactive segmentation (Santner 2010)

Labels	2	3	4	5	6	7	8	9	13
Frequency	66	104	58	18	11	2	2	1	1

Table 2 Comparison of the examined methods for PDE (PdeZGFN and PdeCCP) and MRF (Alpha4, Alpha8, FastPD4, FastPD8) optimization on the Graz benchmark

Optimization	Average Dice-score 2 regions	Average Dice-score	Average runtime
PdeZGFN	0.9397 (± 0.0724)	0.9322 (± 0.0746)	0.49 (± 0.44) sec
PdeCCP	0.9350 (± 0.0803)	0.9314 (± 0.0782)	1.77 (± 2.46) sec
Alpha4	0.9477 (± 0.0533)	0.9358 (± 0.0686)	1.29 (± 1.37) sec
Alpha8	0.9466 (± 0.0550)	0.9372 (± 0.0680)	2.63 (± 2.59) sec
FastPD4	0.9472 (± 0.0533)	0.9360 (± 0.0682)	0.74 (± 0.63) sec
FastPD8	0.9466 (± 0.0550)	0.9366 (± 0.0683)	1.52 (± 1.81) sec

The mean and standard deviation are given for the Dice-score for the (globally optimally solvable) case of two regions and for the whole database, as well as for the runtime per image on the whole database

Bold are the best performances in each column, e.g. highest score in column 1 and 2 and lowest runtime in column 3

H : The Dice-scores follow the same distribution. (30)

The test is based on the comparison of the empirical cumulative distribution function of each Dice-score dataset. This hypothesis could not be rejected for any two optimization approaches on a significance level of 5%. Figure 3 shows the empirical cumulative distribution function of the Dice-score values for all methods. The plot confirms that the overall quality of the results does not depend on the optimization method.

7.1.2 Visual Evaluation

To visually assess the quality of the compared optimization schemes we show results on a few benchmark images in Fig. 4. When inspecting the results of the algorithms we notice only slight quality differences, which confirm the result of the statistical test.

Visual differences between results of continuous and discrete methods are often due to *metrication errors*, which occur in discrete optimization (Fig. 5c). Region boundaries tend to run either horizontally, vertically or diagonally, since the boundary length is either measured with respect to the L_1 norm (4-connectivity) or an approximated L_2 norm (larger connectivity). This is especially true for regions with uncertain data fidelity.

7.2 Ambiguities of the Segmentation Results

For more than two labels, none of the introduced algorithms yield globally optimal results in general. Hence, ambiguities can appear in the results.

For the two graph cut algorithms, ambiguities arise due to the order in which labels α are visited in each outer iteration. The order of traversal influences both the quality and the runtime of the algorithm. For four regions we registered a run-

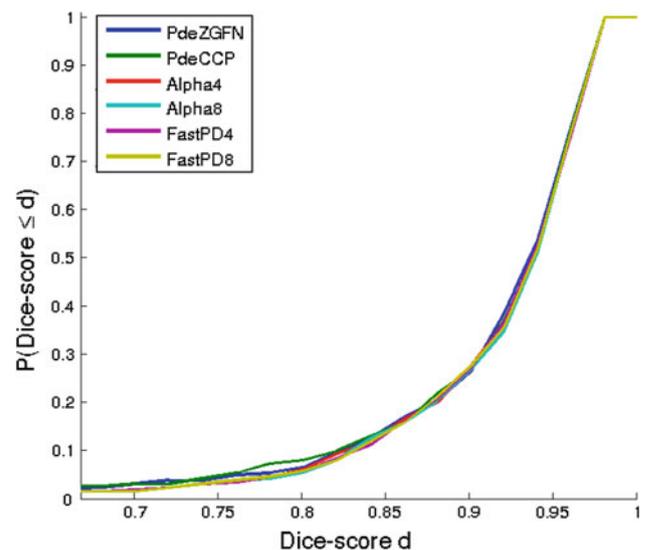


Fig. 3 Statistical evaluation of the Dice-score for all methods. The plot shows the empirical cumulative distribution function of the Dice-score ($P(\text{Dice-score} > d)$) over all benchmark images. The closer to one the function is the higher is the quality of its segmentation results. Based on the Kolmogorov-Smirnov test we could not find any significant quality differences (Color figure online)

time difference of up to five seconds depending on whether we iterated over labels $1, \dots, N$ or $N, \dots, 1$. Segmentation results can also vary locally for different iteration orders, see Fig. 5d. Results also depend on the initialization, which can lead to the flipping of small regions or even whole labels. In Figs. 5f, g the first run was initialized with a constant label, the second with random values in the lower half of the image.

In contrast, with PDEs the order in which the label indicator functions u_i are updated has no impact on the result. This is because each u_i is updated separately, whereas the constraints are enforced at the end of each iteration by projection.



Fig. 4 Results for the compared algorithms on selected images from the Graz benchmark (a) PdeZGFN, (b) PdeCCP, (c) Alpha8, (c) FastPD8

Furthermore, since the relaxed problems 4.3.1 and 4.3.2 are convex, their global minimum will always be attained independent of the initialization. However, ambiguities also occur for PDE-based approaches for more than two labels when the relaxed solution is binarized (14), see Fig. 5e.

7.3 Experimental Optimality Bounds

For both the continuous and the discrete approaches, a priori and a posteriori optimality bounds were given in (16), (21), (15) and (22).

Fig. 5 Ambiguities and metrication errors in optimization results on given input images (a) and (b) from the Graz benchmark. (c) Metrication errors for Alpha4 (top), Alpha8 (center) in contrast to a smooth boundary for the PDE result (bottom), (d) Ambiguous MRF results depending on the traversed label order, (e) Ambiguous PdeCCP results due to binarization, (f,g) Ambiguous MRF results due to different initializations

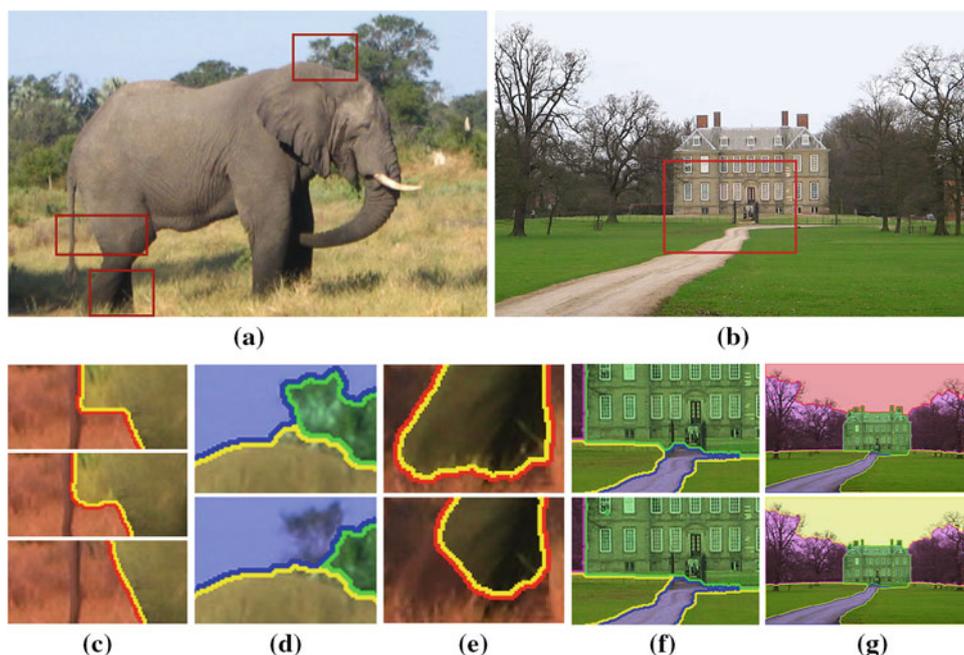


Table 3 Experimental optimality bounds which indicate how far the computed multi-label solutions are from the global minima of the segmentation energies, i.e. the relative gap given in (31)

PdeCCP (%)	PdeZGFN (%)	Alpha4 (%)	FastPD4 (%)
0.18	0.25	0.0002	0.001

The continuous approaches are not comparable to the discrete approaches, but the PdeCCP approach is tighter than the ZGFN approach, and the Alpha4 approach is tighter than the FastPD4 approach

To compute the a posteriori bounds, we use the benchmark results. Let u indicate the integer solution of the respective algorithm and \tilde{u} the result of the solution of the corresponding relaxed optimization problem. Then we can compute the following relative optimality bound

$$G(u, \tilde{u}) = \frac{E(u) - E(\tilde{u})}{E(\tilde{u})}. \quad (31)$$

For computing the relaxed energy of the discrete approaches we optimized the linear program (19). Table 3 shows the average bounds on the Graz benchmark for the considered continuous and discrete approaches. The results suggest that only very small numbers of pixels are not optimal after optimization terminates. Thus, even though the algorithms do not yield the global optimum for more than two labels, they are very close to it with negligible error. Figure 6 shows results of the relaxed optimization problems for the continuous algorithms. The number of non-binary grey pixels without clear label assignment is very small, e.g. the elephant's tooth and the top of the church tower.

Table 3 suggests that the discrete approaches yield even smaller gaps than the continuous approaches. Concluding a higher accuracy of the discrete approaches from these

results would, however, be misleading for two reasons: (1) since the original optimization problems are different (i.e. the discrete approaches only approximate the L_2 -norm), the energy relation between the original problems and their respective relaxations differs as well. Thus the measured bounds are not comparable between continuous and discrete approaches. (2) The error bounds are only upper bounds—there is no indication on how far the optimal binary solution actually is from the computed solution. The results for the continuous approaches confirm that the relaxation proposed by Chambolle et al. is tighter than that by Zach et al. They also suggest that α -expansion results are closer to optimality than Fast-PD results due to different label orders and initializations.

7.4 Runtimes

For interactive image segmentation runtimes close to real-time are indispensable. One of the advantages of PDE based methods is that they can be parallelized and run on graphics hardware. In contrast, computing the maximum flow on a graph in parallel is difficult leading only to limited speedups.

The average runtime for each method on the whole database is indicated in Table 2. The continuous approaches yield lower runtimes on average per image than the discrete methods, with PdeZGFN exhibiting 0.49 s. Among the discrete approaches FastPD4 yields the lowest average runtime with 0.74 s. Note that when PdeCCP is implemented as originally proposed by Chambolle et al. (2008) no exact solution is computed due to the inexact Dykstra projections onto the set (7). However, the results are very similar and the runtime is

Fig. 6 Results of the relaxed optimization problems before binarization (a) by Chambolle et al. and (b) by Zach et al. The fact that the indicator functions are non-binary in very few locations only (marked by red circles) indicates near-optimality of the computed solutions (Color figure online)

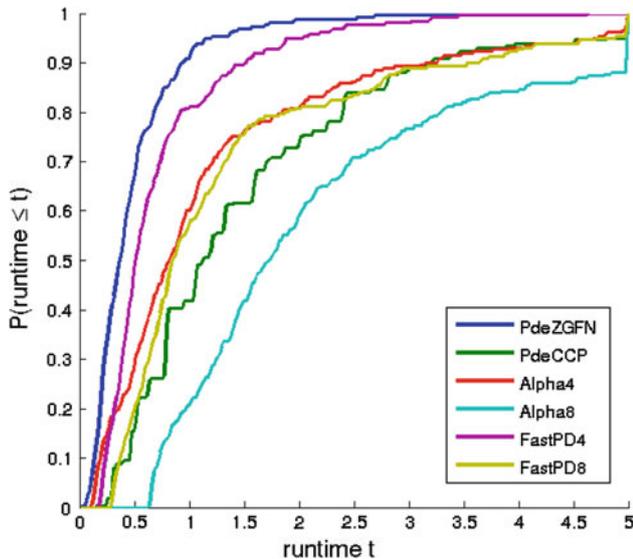
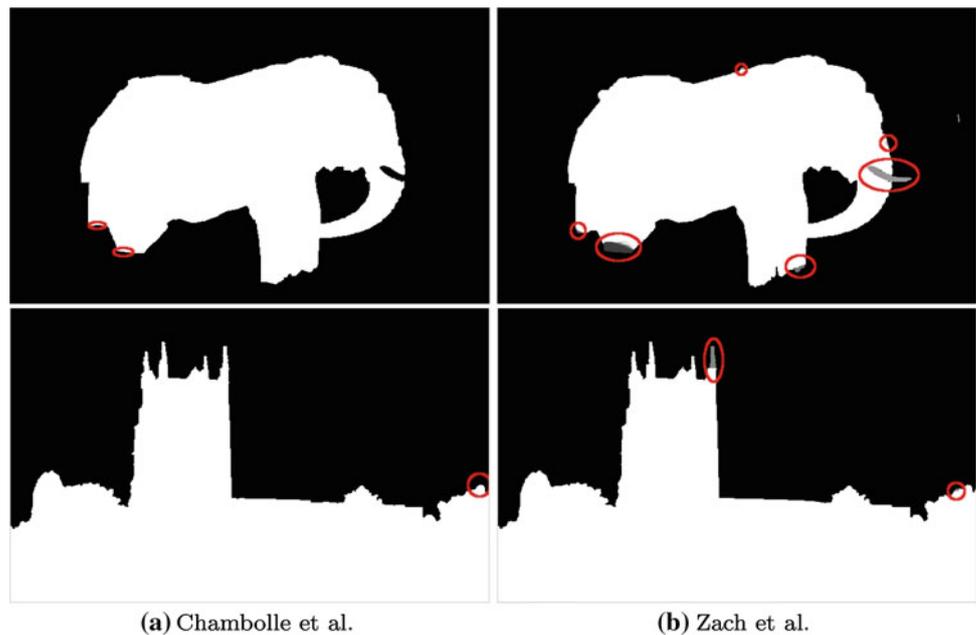


Fig. 7 Statistical evaluation of the runtimes for all methods. The plot shows the cumulative distribution functions of the runtimes ($P(\text{runtime} \leq t)$) over all benchmark images. The closer to one the plotted function is the faster is the corresponding algorithm. The Komogorov-Smirnov test shows that the runtime differences are significant, i.e. PdeZGFN is the fastest algorithm. All runtimes above 5 s are cut to 5 s for better readability (Color figure online)

diminished by a factor of three with strongly reduced memory consumption.

7.4.1 Statistical Evaluation

To statistically evaluate differences in the computational speed between all methods, we used the Kolmogorov-Smirnov

test in order to examine the following hypothesis for each two optimization methods

$$H : \text{The runtimes follow the same distribution.} \quad (32)$$

For the runtimes, this hypothesis was rejected for any two optimization approaches on a significance level of 5% and even of 1%. Figure 7 shows the empirical cumulative distribution functions of the runtimes for all methods. Runtimes above five seconds were cut to the maximum value of five seconds in this Figure to improve readability. The plot confirms that the runtimes strongly differ.

In general, the continuous ZGFN-relaxation exhibits the lowest runtime. Among the continuous methods, only the PdeCCP approach obtains runtimes above five seconds in case of more than six labels, with up to 28.83 s maximum for 13 labels. In contrast, the PdeZGFN approach does not take more than 2.54 s and thus scales well with the number of labels. Comparable results to the PdeZGFN method among the discrete methods can only be obtained by the FastPD4 method. Alpha4 and FastPD8 show similar runtime distributions, whereas Alpha8 yields the longest runtimes on average, however, with only 7.9 s maximum for 13 labels. The percentage of images requiring more than 5 s until convergence are 0% (PdeZGFN), 5% (PdeCCP), 3% (Alpha4), 12% (Alpha8), 0% (FastPD4) and 5% (FastPD8).

7.4.2 Runtimes per Label Number

Finally, we examine the runtimes with respect to the number of labels for all approaches. Figure 8 and corresponding Table 4 show the average runtime in seconds for each label computed over the whole benchmark. For the PDE

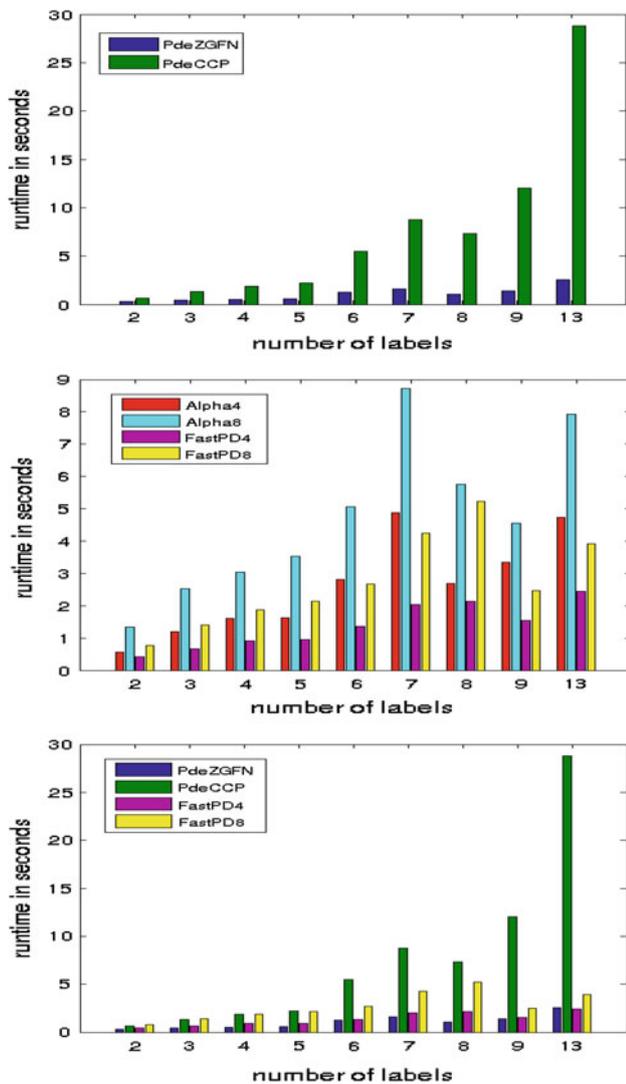


Fig. 8 Comparison of average runtimes for different numbers of labels for PDE optimization schemes (PdeZGFN and PdeCCP) compared to MRF optimization (Alpha4, Alpha8, FastPD4, FastPD8) for the problem of interactive image segmentation

Table 4 Average runtimes in seconds per label number on the Graz benchmark

Method	2	3	4	5	6	7	8	9	13	Avg-std
PdeZGFN	0.29	0.45	0.54	0.56	1.24	1.63	1.04	1.38	2.54	0.51
PdeCCP	0.68	1.33	1.85	2.24	5.50	8.79	7.34	12.03	28.83	2.16
Alpha4	0.56	1.19	1.61	1.63	2.80	4.88	2.70	3.35	4.74	1.03
Alpha8	1.33	2.54	3.03	3.53	5.06	8.72	5.75	4.55	7.93	1.85
FastPD4	0.42	0.67	0.91	0.94	1.35	2.05	2.13	1.54	2.44	0.46
FastPD8	0.77	1.40	1.86	2.14	2.67	4.25	5.22	2.46	3.92	1.22

For the last column we computed the standard deviation of the runtime for each label number (comprising more than a single image) and took the average

Table 5 Average runtimes in seconds of the proposed multi-label segmentation algorithms PdeZGFN and FastPD4 with respect to differently scaled smoothness values

Method	$\lambda_{opt}/100$	$\lambda_{opt}/10$	λ_{opt}	$\lambda_{opt} \cdot 10$	$\lambda_{opt} \cdot 100$
FastPD4	0.37	0.43	0.74	2.10	8.18
PdeZGFN	0.13	0.23	0.49	0.96	1.41

λ_{opt} is the optimal smoothness parameter with respect to the benchmark

approaches the runtime increases with the number of constraints on the dual variables: approximately linearly in the case of the PdeZGFN approach with linearly increasing constraint set (6), approximately quadratically in the case of the PdeCCP approach with quadratically increasing constraint set (7). The runtimes of the MRF approaches are less predictable as they strongly depend on the data term and thus the image content as discussed in Sect. 2. The computation times vary with the smoothness parameter λ as shown exemplarily for FastPD4 and PdeZGFN in Table 5.

8 Conclusion

In this paper we have reviewed and experimentally compared the most popular recent multi-label relaxations and optimization methods for the Potts model from the continuous PDE and discrete MRF domain. To evaluate their performance with respect to the quality of the results and runtimes, we applied them to the problem of interactive multi-label segmentation. We statistically tested the hypotheses that the quality and runtime followed the same distribution for each two of the optimization methods under consideration. The quality hypothesis could not be rejected, whereas the runtime hypothesis was rejected. We conclude that all optimization approaches yield results of comparable quality, whereas the runtimes differ. The lowest runtime could be achieved with the relaxation by Zach et al. with 0.49 s on average per image containing 2–13 labels.

Acknowledgments We thank Vladimir Kolmogorov and Evgeny Strelakovsky for fruitful discussions on the relations among various relaxations.

References

- Alahari, K., Kohli, P., & Torr, P. H. S. (2010). Dynamic hybrid algorithms for MAP inference in discrete MRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10), 1846–1857.
- Batra, D. (2011). Making the right moves: Guiding alpha-expansion using local primal-dual gaps. In P. Kohli (Ed.), *International Conference on Computer Vision and Pattern Recognition*, Colorado Springs.
- Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of Royal Statistical Society Series B*, 48(3), 259–302.
- Boykov, Y., & Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9), 1124–1137.
- Boykov, Y., Veksler, O., & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11), 1222–1239.
- Chambolle, A., Cremers, D., Pock, T. (2008). A convex approach for computing minimal partitions. Technical Report TR-2008-05. Bonn: University of Bonn.
- Chan, T., & Esedoglu, S., & Nikolova, M., (2006). Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5), 1632–1648.
- Chan, T., & Vese, L. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10(2), 266–277.
- Cremers, D., Rousson, M., & Deriche, R. (2007). A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *International Journal of Computer Vision*, 72(2), 195–215.
- Cremers, D., Sochen, N., & Schnörr, C. (2004). Multiphase dynamic labeling for variational recognition-driven image segmentation. In T. Pajdla & V. Hlavac (Eds.), *European Conference on Computer Vision volume of 3024 LNCS* (pp. 74–86). New York: Springer.
- Cremers, D., Sochen, N., & Schnörr, C. (2006). A multiphase dynamic labeling model for variational recognition-driven image segmentation. *International Journal of Computer Vision*, 66(1), 67–81.
- Felzenszwalb, P., & Veksler, O. (2010). Tiered scene labeling with dynamic programming. In *International Conference on Computer Vision and Pattern Recognition*, San Francisco.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 721–741.
- Goldschlager, L., Shaw, R., & Staples, J. (1982). The maximum flow problem is log space complete for P. *Theoretical Computer Science*, 21, 105–111.
- Greig, D. M., Porteous, B. T., & Seheult, A. H. (1989). Exact maximum a posteriori estimation for binary images. *Journal of Royal Statistical Society Series B*, 51(2), 271–279.
- Ishikawa, H. (2003). Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10), 1333–1336.
- Kleinberg, J., & Tardos, E. (2002). Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *Journal of the ACM*, 49(5), 672–713.
- Klodt, M., Schoenemann, T., Kolev, K., Schikora, M., & Cremers, D. (2008). An experimental comparison of discrete and continuous shape optimization methods. In *European Conference on Computer Vision*, Marseille, France.
- Kolev, K., Pock, T., & Cremers, D. (2010). Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. In *European Conference on Computer Vision*, Crete.
- Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, 1568–1583.
- Komodakis, N., & Tziritas, G. (2005). A new framework for approximate labeling via graph cuts. In *IEEE International Conference on Computer Vision*, New Orleans.
- Komodakis, N., Tziritas, G., & Paragios, N. (2007). Fast, approximately optimal solutions for single and dynamic MRFs. In *International Conference on Computer Vision and Pattern Recognition*, Ezhou.
- Lellmann, J., Becker, F., & Schnörr, C. (2009). Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In *IEEE International Conference on Computer Vision* (pp. 646–653).
- Lellmann, J., Kappes, J. H., Yuan, J., Becker, F., & Schnörr, C. (2009). Convex multi-class image labeling by simplex-constrained total variation. *Scale Space and Variational Methods in Computer Vision (SSVM)*, 5567, 150–162.
- Lellmann, J., Lenzen, F., & Schnörr, C. (2011). Optimality bounds for a variational relaxation of the image partitioning problem. In *International Conference on Energy Minimization Methods for Computer Vision and Pattern Recognition*. New York: Springer.
- Lempitsky, V., Rother, C., & Blake, A. (2007). Logcut: Efficient graph cut optimization for Markov random fields. In *IEEE International Conference on Computer Vision*.
- Liu, X., Veksler, O., & Samarabandu, J. (2010). Order preserving moves for graph cut based optimization. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 32(7), 1317–1324.
- Michelot, C. (1986). A finite algorithm for finding the projection of a point onto the canonical simplex of R^n . *Journal of Optimization Theory and Applications*, 50(1), 189–193.
- Mumford, D., & Shah, J. (1989). Optimal approximation by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42, 577–685.
- Nieuwenhuis, C., & Cremers, D. (2012). Spatially varying color distributions for interactive multi-label segmentation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Philadelphia.
- Nieuwenhuis, C., & Töppe, E., & Cremers, D. (2011). Space-varying color distributions for interactive multiregion segmentation: Discrete versus continuous approaches. In *International Conference on Energy Minimization Methods for Computer Vision and Pattern Recognition*, New York.
- Osokin, A., Vetrov, D., & Kolmogorov, V. (2011). Submodular decomposition framework for inference in associative markov networks with global constraints. In *International Conference on Computer Vision and Pattern Recognition*, St. Petersburg.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufmann.
- Pock, T., & Chambolle, A. (2011). Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *IEEE International Conference on Computer Vision*, Barcelona.
- Pock, T., Cremers, D., Bischof, H., & Chambolle, A. (2009). An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In *IEEE International Conference on Computer Vision*, Kyoto.
- Pock, T., Cremers, D., Bischof, H., & Chambolle, A. (2010). Global solutions of variational models with convex regularization. *SIAM Journal of Imaging Sciences*, 3(4), 1122–1145.
- Santner, J. (2010). *Interactive multi-label segmentation*. Ph.D. thesis, University of Graz, Graz.
- Schlesinger, M. I. (1976). Sintaksicheskiy analiz dvumernykh zritel'nikh signalov v usloviyakh pomekh (syntactic analysis of two-

- dimensional visual signals in noisy conditions). *Kibernetika*, 4, 113–130. (in Russian).
- Stekalovskiy, E., & Cremers, D. (2011). Generalized ordering constraints for multilabel optimization. In *IEEE International Conference on Computer Vision*, Barcelona.
- Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., & Rother, C. (2006). A comparative study of energy minimization methods for Markov random fields. In *European Conference on Computer Vision*, volume 3952 of *Lecture Notes in Computer Science*, Graz (pp. 16–29).
- Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., & Fan, A., et al. (2001). Model-based curve evolution technique for image segmentation. In *Computer Vision Pattern Recognition, Kauai, Hawaii* (pp. 463–468).
- Veksler, O. (2007). Graph cut based optimization for MRFs with truncated convex priors. In *International Conference on Computer Vision and Pattern Recognition*, Beijing.
- Veksler, O. (2009). Multi-label moves for MRFs with truncated convex priors. In *International Conference on Energy Minimization Methods for Computer Vision and Pattern Recognition*, Bonn.
- Wainwright, M., Jaakkola, T., & Willsky, A. (2005). Map estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches. *IEEE Transactions on Information Theory*, 51, 3697–3717.
- Werner, T. (2007). A linear programming approach to maxsum problem: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7), 1165–1179.
- Zach, C., Gallup, D., Frahm, J. M., & Niethammer, M. (2008). Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision modeling and visualization workshop (VMV)*, Konstanz.
- Zach, C., Häne, C., & Pollefeys, M. (2012). What is optimized in tight convex relaxations for multi-label problems? In *International Conference on Computer Vision and Pattern Recognition*, Lund Sweden.
- Zach, C., Niethammer, M., & Frahm, J. M. (2009). Continuous maximal flows and Wulff shapes: Application to MRFs. In *International Conference on Computer Vision and Pattern Recognition*, Bonn.