

Space-Varying Color Distributions for Interactive Multiregion Segmentation: Discrete versus Continuous Approaches

Claudia Nieuwenhuis, Eno Töppe and Daniel Cremers

Department of Computer Science, TU München, Germany

Abstract. State-of-the-art approaches in interactive image segmentation often fail for objects exhibiting complex color variability, similar colors or difficult lighting conditions. The reason is that they treat the given user information as independent and identically distributed in the input space yielding a single color distribution per region. Due to their strong overlap segmentation often fails. By statistically taking into account the local distribution of the scribbles we obtain spatially varying color distributions, which are locally separable and allow for weaker regularization assumptions. Starting from a Bayesian formulation for image segmentation, we derive a variational framework for multi-region segmentation, which incorporates spatially adaptive probability density functions. Minimization is done by three different optimization methods from the MRF and PDE community. We discuss advantages and drawbacks of respective algorithms and compare them experimentally in terms of segmentation accuracy, quantitative performance on the Graz benchmark and speed.

1 Introduction

Segmentation denotes the task of dividing an image into meaningful, non-overlapping regions. Meaningful, especially in complex images, depends on the user's intention of what he wants to extract from the image. This makes the problem highly ill-posed, so user interaction is indispensable. Typically bounding boxes, contours or scribbles are used to indicate the regions of interest. Such interactive segmentation algorithms are widely used in image editing software packages, e.g. for the identification of specific structures in medical images, for tracking or to compose scenes from different images.

Previous interactive approaches use various methods to estimate color models for each object to be segmented. In [21] the authors compute color histograms and threshold histogram distances, in [17] Mixtures of Gaussians are employed to estimate the color distributions, in [3] kernel density estimators are used to estimate geodesic distances between each two pixels, and in [19] methods from machine learning have been applied to the task.

All of these approaches derive a single, constant color model for each region. This policy does not adequately represent the given user information, since

scribbles contain information on color and space. By disregarding the spatial information in kernel density estimation we assume that the color information is, in a statistical sense, spatially independent and identically distributed (iid) over the whole image. This is not correct and only leads to good results in case of a strong, well-chosen prior and non-overlapping color distributions - which is, clearly, not the case in natural images with changing light conditions, multicolored objects and difficult textures. In fact, close to a scribble we have precise information on the color of the object, at scribble pixels we even know the color exactly. Only if we are far away from the scribbles spatial location becomes less important, since we have no precise knowledge on the object color. This is not reflected by previous data fidelity terms. In [13] we segmented optical flow fields by means of spatially dependent flow vector distributions. The first contribution of this paper is to show how non-iid assumptions can be handled in color density estimation leading to strongly improved results.

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. In the discrete MRF community the segmentation problem is related to the classical Potts model [15], in the spatially continuous setting it is typically referred to as the minimal partition problem. Over the years a number of algorithms for multilabel optimization have been developed, both in the MRF community and in the PDE community. As a result one may ask how respective algorithms compare in practice. In [9] an experimental comparison of discrete and continuous optimization approaches was presented for the specific case of *binary* labeling problems. In recent years, the focus has shifted from binary labeling problems to the more general multilabel problem. The second contribution of this paper is to perform an experimental comparison of respective algorithms for the minimal partition problem with more than two regions or labels. Such comparisons are invariably limited as they can only reflect the state of the art for a given time whereas algorithms are continuously being improved. In this paper, we will therefore focus on an experimental comparison of the following algorithms, which minimize the same energy in a discrete and in a continuous formulation:

- The discrete alpha expansion approach with 4 and 8 connectivity [5].
- The continuous approach by Chambolle et al. [6]
- The continuous approach by Lellmann et al. [12]

Yet, we should point out that even at this moment, there exist further accelerations of algorithms such as primal-dual algorithms [10] and dynamic hybrid algorithms [2] for speeding up MRF optimization and Douglas-Rachford splitting which reportedly speeds up the Chambolle relaxation by a factor of 4-20 [11]. We qualitatively and quantitatively compare results on the specific problem of multilabel image segmentation based on a known segmentation benchmark.

2 A Statistical Framework for Segmentation

Let $I : \Omega \rightarrow \mathbb{R}^d$ denote the input image defined on the domain $\Omega \subset \mathbb{R}^2$ (in the discrete case $\Omega = \{1..N\} \times \{1..M\}$). The task of segmenting the image plane into a set of n pairwise disjoint regions Ω_i

$$\Omega = \bigcup_{i=1}^n \Omega_i, \quad \Omega_i \cap \Omega_j = \emptyset \quad \forall i \neq j \quad (1)$$

can be solved by computing a labeling $l : \Omega \rightarrow \{1, \dots, n\}$ indicating which of the n regions each pixel belongs to: $\Omega_i = \{x | l(x) = i\}$.

For multilabel segmentation the energies to be minimized usually contain two terms: a data term $\rho_i : \Omega \rightarrow \mathbb{R}$ indicating how well the observed data fits to region $i \in \{1, \dots, n\}$, and a length regularization term.

2.1 Segmentation as Bayesian Inference

In the framework of Bayesian inference, one can compute a segmentation by maximizing the conditional probability

$$\arg \max_l \mathcal{P}(l | I) = \arg \max_l \mathcal{P}(I | l) \mathcal{P}(l). \quad (2)$$

Assuming that the colors of all pixels are independent of each other, but – in contrast to previous interactive segmentation approaches – not independent of space, we obtain

$$\mathcal{P}(I | l) = \left(\prod_{x \in \Omega} \left(\mathcal{P}(I(x) | x, l) \right)^{dx} \right), \quad (3)$$

where the exponent dx denotes an infinitesimal volume in \mathbb{R}^2 and assures the correct continuum limit. Note how the space-dependency of color likelihoods arises naturally in this formulation. It has commonly been neglected, yet we shall show in this paper that taking into account this spatial variation of color distributions based on scribble locations leads to drastic improvements of the resulting interactive segmentation process. Assuming furthermore that the color probability at location x does not depend on the labeling of other pixels $y \neq x$, the product in (3) can be written as

$$\mathcal{P}(I | l) = \prod_{i=1}^n \prod_{x \in \Omega_i} \left(\mathcal{P}(I(x) | x, l(x) = i) \right)^{dx}. \quad (4)$$

2.2 Inferring Space-Variant Color Distributions

The expression $\mathcal{P}(I(x) | x, l(x) = i)$ in (4) denotes the conditional probability for observing a color value I at location x given that x is part of region Ω_i . It can be estimated from the user scribbles as follows. Let

$$S_i := \left\{ \begin{pmatrix} x_{ij} \\ I_{ij} \end{pmatrix}, j = 1, \dots, m_i \right\} \quad (5)$$

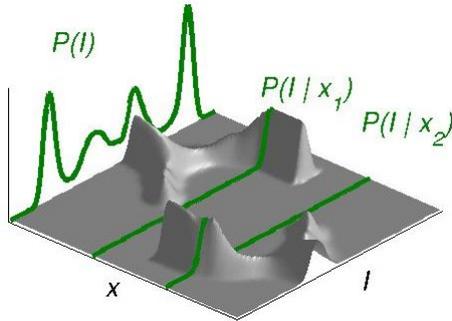


Fig. 1. Schematic plot of a kernel density estimate in the joint space of location x and color I computed for a set of sample points. If space is disregarded we obtain the same density function $P(I)$ at each pixel showing several peaks for colors, which are predominant in different object parts. Normalizing the joint distribution gives the conditional probability $\mathcal{P}(I|x)$ for observing a color I at a location x we obtain a separate color distribution showing only a single color peak for the locally predominant color. This distribution is used as a data term for image segmentation.

denote the set of user scribbles S_i for region i consisting of pixel locations x_{ij} with corresponding color values I_{ij} . The probability of a given color value I at location x within region i is computed by a Parzen density estimator [1, 16] of the form:

$$\hat{\mathcal{P}}(I, x | l(x) = i) = \frac{1}{m_i} \sum_{j=1}^{m_i} k \left(\begin{array}{c} x - x_{ij} \\ I - I_{ij} \end{array} \right), \quad (6)$$

i.e. a sum of normalized kernel functions k centered at each sample point in the product space of location and color.

Figure 1 shows a schematic drawing of this distribution in the joint space of color and spatial coordinate estimated from a set of sample points. Commonly in interactive segmentation the location of the scribbles is not taken into account and the space-independent color distribution, the marginal

$$\hat{\mathcal{P}}(I | l(x) = i) = \int \hat{\mathcal{P}}(I, x | l(x) = i) dx, \quad (7)$$

is used to estimate color likelihoods. The marginal is plotted on the left and contains four peaks, each one for a different predominant color of the foreground object. At each location in the image, the likelihood for each of these colors follows the same marginal distribution, no matter if we are very close to a scribble or far away. In this paper, instead of the marginal the joint distribution of color and space is estimated to account for color changes in segmentation regions. Typically, the color is not independent of space. In this way, one obtains a separate color distribution at each location in the image.

Unfortunately, in the considered scenario the Parzen density estimator (6) is not guaranteed to converge to the true density since the samples are not independent

and identically distributed (iid), but rather user-placed in locations which are by no means independent. To account for this dependency we employ spatially adaptive kernels. In practice, we choose isotropic Gaussian kernels with variance σ and ρ in color and space dimension. Due to the separability of the Gaussian kernel we can write:

$$k \begin{pmatrix} x \\ I \end{pmatrix} = k_{\rho(x)}(x) k_{\sigma}(I). \quad (8)$$

To account for the non-uniform spatial distribution, the spatial kernel variance $\rho(x)$ depends on the image location x and is chosen proportional to the distance from the k -th nearest scribble point $x_{v_k} \in S_i$: $\rho(x) = \alpha|x - x_{v_k}|_2$.

3 Discrete versus Continuous Energies

There are two basic approaches for the optimization of multilabel problems: MRF approaches and PDE approaches. In the following we will review both paradigms and point out practical advantages and drawbacks.

3.1 MRF approaches

For solving multilabel segmentation in the case of a discrete image domain we seek to minimize the following energy

$$E(l) = \sum_{i \in \Omega} D_i(l(i)) + \sum_{(i,j) \in \mathcal{N}} S_{ij}(l(i), l(j)) \quad (9)$$

where $D_i : \{1, \dots, n\} \rightarrow \mathbb{R}$ measures the data fidelity (see section 2) and $S_{ij} : \{1, \dots, n\} \times \{1, \dots, n\} \rightarrow \mathbb{R}$ the smoothness between pixels i and j . The smoothness is summed over all pixel pairs in the neighborhood \mathcal{N} . In the simplest case, \mathcal{N} consists of the upper, lower, left and right neighbor of each pixel. This is called *4-connectivity*. If the diagonal neighbors belong to \mathcal{N} as well, this is called *8-connectivity*. In discrete approaches, the image is represented by a graph containing one node for each pixel and links between neighboring nodes, whose weight is associated with the smoothness term. This constitutes a Markov Random Field interpretation of the problem (9).

In [20] different algorithms for the minimization of Markov Random Field approaches are compared: iterated conditional modes, graph cuts (alpha expansion and alpha-beta-swaps) and message passing algorithms such as belief propagation. Of those algorithms, alpha expansion yields the best performance results for the Pott's model, which is a suitable model for the multiregion segmentation problem. Hence, in this paper we use the graph cut alpha-expansion approach [5] for optimization in the discrete case.

The basic idea of graph cuts 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 crossed by the partition interface should be minimal of all possible separations. One can show that this problem is equivalent to computing a maximum flow which can be done in polynomial time.

In computer vision, graph cuts were applied for the first time by Greig, Porteous and Seheult [8]. They showed that for the problem of denoising a *binary* image a global optimum of the corresponding MAP formulation can be found by computing the maximum flow through a certain graph construction. Later Boykov et al. revisited the idea exploring to a greater extent which energies can be minimized by graph cuts [5].

If the smoothness penalty constitutes a metric on the label space, an iterative method called *alpha expansion* can be applied. It was shown in [5] that the computed result is energetically within a factor of two of the global optimum. In the alpha expansion algorithm the multilabel problem is reduced to a number of binary labeling problems which are iterated. The basic idea is that in each step for a selected label α pixels can either keep their current label or take the label α . To this end, a special graph is constructed and the min cut is computed. The steps are executed for each label α in the label set until convergence. The binary labeling problem of each step is solved globally optimal, the solution to the multilabel segmentation problem, however, is only locally optimal. In this paper we refer to the alpha expansion algorithm based on the interactive segmentation dataterm with 4-connectivity as *Gc4*, with 8-connectivity as *Gc8*. It is based on the max-flow implementation by Boykov and Kolmogorov [4].

3.2 PDE Approaches

In the continuous setting the optimization problem is formulated as a set of partial differential equations (PDE) based on region indicator functions. The energy in the continuous setting also consists of the dataterm and a length regularization term. In case of convex energies or relaxations the resulting problem can be solved by fast primal-dual optimization schemes.

For the multilabel minimal partition problem (the Pott's model) three main PDE based approaches have been proposed in parallel in recent years, e.g. the approach by Chambolle et al. [6], Zach et al. [22] and by Lellmann et al. [12]. These approaches differ in their complexity and by the tightness of the energy relaxation. None of these approaches is convex, but convex relaxations can be formulated to tackle the optimization. However, as in the discrete case these approaches only lead to locally optimal solutions.

Lellmann et al. [12] represent the n regions Ω_i by the indicator function $u \in \text{BV}(\Omega, \{0, 1\})^n$.

$$u_i(x) = \begin{cases} 1, & \text{if } x \in \Omega_i \\ 0, & \text{else} \end{cases} \quad \forall i = 1, \dots, n. \quad (10)$$

Here BV denotes the function space of bounded variation. Based on this definition they formulated the following energy given the dataterm ρ_i

$$E(u) = \sum_{i=1}^n \int_{\Omega} u_i \rho_i + \lambda \sum_{i=1}^n \int_{\Omega} g(x) |Du_i| \, dx \quad \text{s.t.} \quad \sum_{i=1}^n u_i(x) = 1, \quad x \in \Omega, \quad (11)$$

which is minimized. The first term punishes incoherence with the observed data, whereas the second term measures the length of the region boundaries. Here Du_i

denotes the distributional gradient of u_i , λ balances the data term and the length regularization and $g(x) = \exp(-\gamma|\nabla I(x)|)$ favors the coincidence of object and image edges. To handle the non-differentiable indicator functions, the boundary of the set indicated by u_i , the perimeter, can be written as the total variation:

$$\text{Per}(\{x \mid u_i(x) = 1\}) = \int_{\Omega} g(x) |Du_i| = \sup_{\xi: |\xi(x)| \leq g(x)} \left(- \int_{\Omega} u_i \text{div } \xi_i \, dx \right), \quad (12)$$

where $\xi_i \in C_c^1(\Omega, \mathbb{R}^2)$ denote the dual variables and C_c^1 the space of smooth functions with compact support. With this notation, the original energy minimization problem (11) can be relaxed to

$$\min_{u \in \mathcal{B}} \sup_{\xi \in \mathcal{K}_1} \left\{ \int_{\Omega} u_i \rho_i \, dx - \lambda \sum_{i=1}^n \int_{\Omega} u_i \text{div } \xi_i \, dx \right\} \quad (13)$$

$$\mathcal{B} = \left\{ u \in \text{BV}(\Omega, [0, 1]^n) \mid \sum_{i=1}^n u_i = 1 \right\},$$

$$\mathcal{K}_1 = \left\{ \xi \in C_c^1(\Omega, \mathbb{R}^{2 \times n}) \mid \|\xi(x)\|_F \leq g(x), x \in \Omega \right\}, \quad (14)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. For multilabel segmentation we compute the interactive segmentation data term described in the previous section. We will refer to the resulting approach based on the optimization scheme by Lellmann et al. as *PdeL*.

Chambolle et al. [6] define the following region indicator function

$$u_i(x) = \begin{cases} 1, & \text{if } l(x) \geq i \\ 0, & \text{else} \end{cases} \quad \forall i = 1, \dots, n. \quad (15)$$

with $1 = u_0 \geq u_1(x) \geq \dots \geq u_n(x) \geq u_{n+1} = 0$, where $u_0 = 1$ and $u_{n+1} = 0$ are added to simplify notation. Based on this indicator function the following energy is proposed:

$$E(u) = \sum_{i=1}^n \int_{\Omega} (u_i - u_{i+1}) \rho_i + \lambda \sum_{i=1}^n \int_{\Omega} g(x) |Du_i| \, dx \quad (16)$$

$$1 = u_0 \geq u_1(x) \geq \dots \geq u_n(x) \geq u_{n+1} = 0. \quad (17)$$

Since this energy is not convex due to the binary constraints on u Chambolle et al. propose the following relaxation

$$\min_{u \in \mathcal{B}} \sup_{\substack{\xi \in \mathcal{K}_2 \\ \eta \in \mathcal{W}}} \sum_{i=1}^n \int_{\Omega} (u_i - u_{i+1}) \eta_i - \lambda \sum_{i=1}^n \int_{\Omega} u_i \text{div } \xi_i \, dx \quad (18)$$

$$\mathcal{K}_2 = \left\{ \xi : \Omega \rightarrow \mathbb{R}^{2 \times n} \mid \left| \sum_{i_1 \leq i \leq i_2} \xi_i \right| \leq g(x), 1 \leq i_1 \leq i_2 \leq n \right\} \quad (19)$$

$$\mathcal{W} = \left\{ \eta : \Omega \rightarrow \mathbb{R}^n \mid |\eta_i(x)| \leq \rho_i(x), x \in \Omega, 1 \leq i \leq n \right\}$$

We will refer to this optimization method applied to the problem of multilabel segmentation based on the interactive segmentation dataterm as *PdeC*.

Both energy formulations are relaxed to obtain convex approaches, which can be globally minimized by primal-dual optimization schemes. In case of two regions the thresholding theorem applies saying that the thresholded globally optimal solution of the relaxed problem yields the globally optimal solution of the original binary problem independent of the chosen threshold. In the multilabel case the threshold theorem does not hold. However, the binarized solution of the relaxed approach has been experimentally shown to lie within very tight bounds from the optimal binary solution. Hence, the proposed relaxations lead to very near optimal solutions of the original problem.

The approach by Lellmann et al. is simpler than the one by Chambolle et al. since its relaxation leads to simpler constraints on the dual variables ξ . Besides it is faster because the number of constraints on the dual variables increases only linearly with the number of labels. In contrast, the approach by Chambolle et al. demands quadratically many constraints on the dual variables yielding higher computation times. However, in [14] the relaxation proposed by Chambolle et al. has been shown to yield tighter bounds meaning that the solution of the relaxed problem is closer to the global optimum of the binary problem than the relaxed result of the approach by Lellmann et al..

3.3 Advantages and Disadvantages

In applications of MRF approaches and PDEs we encountered several advantages and short-comings many of which are known to researchers working on these topics. In particular:

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 the choice of appropriate step size parameters and termination criterion. While appropriate parameters can be determined automatically by convergence analysis and the primal-dual gap, the latter is often not straight-forward to compute.

The commonly used alpha expansion for solving general multilabel MRFs is based on iteratively solving binary problems. As we show in Figure 3 results may therefore depend on the order of the labels. Furthermore graph cut approaches exhibit metrication errors due to a rather crude approximation of the Euclidean boundary length – see Section 4.4. Instead the PDE approaches for multilabel optimization are based on minimizing a single convex functional. They provide smooth object boundaries that do not exhibit a prominent grid bias. Furthermore, 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 [7]. Instead, the considered PDE approaches are easily parallelizable yielding drastic speedups over CPU algorithms.



Fig. 2. Results for the 4 compared algorithms on selected images from the Graz benchmark. a) Santner’s result with indicated user scribbles, b) PdeC, c) PdeL, d) Gc4.

4 Comparison of Segmentation Accuracy

In this section we compare results of the proposed multilabel segmentation approaches *PdeL*, *PdeC* and *Gc4*. To compute the interactive dataterm from user scribbles the following parameters were used: $\sigma = 1.2$, $\alpha = 0.7$, $\gamma = 5$.

For the PDE approaches we set $\lambda = 67$ and the step size $\tau = 0.28$. We used a GTX 580 for the parallel computations. For the alpha expansion we used $\lambda = 6$. Computations were carried out on an Intel Core i7 CPU 860, 2.80GHz.

4.1 Results on the Graz Benchmark

For automatic segmentation several benchmarks are available, e.g. the Berkeley database, the GrabCut database or the Pascal VOC Database. As extensively discussed in [18], these benchmarks are not suited for testing interactive segmentation. Hence, Santner et al. recently published the first benchmark for interactive scribble based multilabel segmentation containing 262 seed-groundtruth pairs from 158 natural images containing between 2 and 13 user labeled segments. The label frequencies are not uniformly distributed. Instead, small numbers such

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

Table 1. Appearance frequency for different label numbers in Santner’s database.

as 2, 3 or 4 labels appear frequently, whereas large numbers are rare (see Table 1). This influences the average quality and runtime values we compute later. To evaluate the segmentation accuracy Santner et al. compute the arithmetic mean of the Dice-score over all segments. It relates the overlap area of the groundtruth $\bar{\Omega}_i$ and the computed segment Ω_i to the sum of their areas

$$dice(\bar{\Omega}_i, \Omega_i) = \frac{2|\bar{\Omega}_i \cap \Omega_i|}{|\bar{\Omega}_i| + |\Omega_i|}. \quad (20)$$

The closer to 1 the Dice-score the more accurate is the segmentation. To evaluate the proposed spatially varying dataterm and the quality of the three different optimization schemes we compute the average Dice-score on the benchmark, which is given in Table 4.1, and compare the results to Santner’s approach.

Santner et al. [19] show impressive results for different combinations of color and texture features in a random forest approach: RGB, HSV and CIELab colors combined with image patches, Haralick features and Local Binary Patterns (LBP). However, they neglect the locality of the scribbles by estimating a single, invariant color model for each region. In our experiments we tested the proposed approach with spatially constant and spatially varying color models on their benchmark. If we use the spatially constant model the results are comparable to those obtained by Santner et al. (RGB color information without texture). They obtain the best results combining CIELab and LBP features in a 21 dimensional vector based on a scribble brush of radius 13. We obtain better results on the benchmark by allowing for spatially varying color models. To obtain the spatially constant approach we set α to a very large value, for the space-variant approach we use $\alpha = 0.7$ to obtain locally adaptive color distributions. The results summarized in Table 4.1 indicate that merely regarding the spatial location of scribbles provides stronger performance improvements than a multitude of sophisticated features. The proposed approach outperforms Santner’s spatially invariant dataterm with all four optimization schemes. The graph cut approaches (Gc4 and Gc8) yield qualitatively better results than Santner’s approach, especially in case of 8-connectivity. The PDE-based approaches yield results comparable to the graph cuts or slightly better. The relaxation proposed by Lellmann et al. (PdeL) is less tight than that by Chambolle et al. (PdeC) leading to results slightly inferior to those by Chambolle et al.

4.2 Visual Results

To visually assess the quality of the three compared optimization schemes we show results on a few benchmark images in Figure 2. When inspecting the results of the three algorithms we notice only slight quality differences, e.g. in the foot of the elephant.

| Method | Dim | Brush | Optim. | Dice-Score |
|----------------------------------|-----|-------|--------|------------|
| [19], RGB | 3 | - | RF | 0.877 |
| our approach, spatially constant | 3 | 3 | PdeL | 0.872 |
| [19], CIELab + LBP | 21 | 5 | RF | 0.917 |
| our approach, space-variant | 3 | 5 | PdeL | 0.923 |
| [19], CIELab + LBP | 21 | 13 | RF | 0.927 |
| our approach, space-variant | 3 | 13 | Gc4 | 0.929 |
| our approach, space-variant | 3 | 13 | Gc8 | 0.931 |
| our approach, space-variant | 3 | 13 | PdeL | 0.931 |
| our approach, space-variant | 3 | 13 | PdeC | 0.934 |

Table 2. Comparison of the proposed spatially varying color model with four different optimization schemes based on random forest (RF), graph cuts (Gc4, Gc8) and PDEs (PdeL, PdeC) to spatially constant color models on the Graz Benchmark. For each approach the dimension of the scribble brush and the average Dice-score is indicated.

4.3 Ambiguities

The alpha expansion algorithm as well as the PDE-based approaches do not lead to globally optimal solutions. Hence, ambiguities can appear in the results. The alpha expansion algorithm iterates over all possible labels α and each time solves a binary graph cut problem. This process is repeated until convergence. The order of traversal influences both the quality and the runtime of the algorithm. For four regions we registered a runtime difference of up to five seconds depending on whether we iterated over labels in order $1..N$ or $N..1$ in each step of the algorithm. Results can differ locally depending on the iteration order as well. An example is shown in Figure 3 b).

In contrast, when solving PDEs each indicator function u_i is updated separately and the variable constraints are only enforced at the end of each iteration, so the iteration order has no impact on the result. Furthermore, since the relaxed problems (13) and (18) are convex, we will always reach their global minimum independent of the label order.

However, ambiguities also occur for PDE-based approaches in the multilabel case. The relaxed problems are convex and thus lead to globally optimal solutions. But to obtain a segmentation the solution must be transformed into a binary result. In case of more than two regions this transformation is not uniquely defined and can be carried out e.g. by thresholding or maximization over the indicator functions. This leads to ambiguous results depending on the chosen method (see Figure 3).

4.4 Metrication Errors

In discrete optimization metrication errors can occur (Figure 3 a). Region boundaries tend to run either horizontally, vertically or diagonally, since the boundary length is not equally penalized in different directions, e.g. in the sea gull image in Figure 2. This is especially true for regions with uncertain data fidelity. The

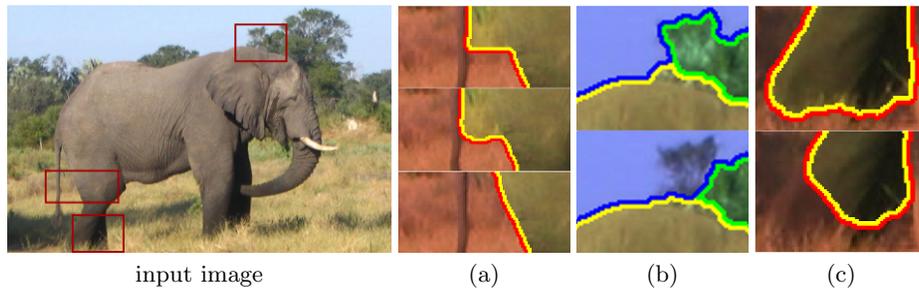


Fig. 3. Metrication errors and ambiguities in optimization results on a given input image. a) Metrication error for Gc4 (top), Gc8 (center) in contrast to a smooth boundary for the PDE result (bottom), b) Ambiguous graph cut results depending on the traversed label order, c) Ambiguous *PdeC* results due to thresholding.

rotationally invariant L_2 norm which is optimized in the continuous case is approximated in the discrete case. With a 4-connectivity boundary the length is penalized with respect to the L_1 norm. The larger the neighborhood, the closer comes the penalization to the L_2 norm at the cost of very large graphs and computation times as well as memory consumption.

5 Runtimes

Especially for interactive image segmentation runtimes close to real-time are indispensable. One of the great 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 with augmenting path algorithms is difficult and only leads to limited speedups. For the PDE approaches the runtime increases with the number of constraints on the dual variables: linearly in case of the PdeL approach (14), quadratically in case of the PdeC approach (19). Figure 4 shows the average runtime in seconds for each label computed over the whole database. The average runtime for the PdeL-method is 0.43 seconds in contrast to 1.27 seconds for the PdeC-method, which is a factor of three higher.

For alpha expansion the runtimes seem to increase linearly in the number of labels. However, the average runtime on the whole database is 1.24 seconds for 4-connectivity and 2.54 seconds for 8-connectivity, which exceeds the PDE approaches by a factor of 2.9 and 5.9 respectively. These computation times vary with the smoothness λ as shown exemplarily for Gc4 and PdeL in Table 5.

| Method | $\lambda_{opt}/100$ | $\lambda_{opt}/10$ | λ_{opt} | $\lambda_{opt} \cdot 10$ | $\lambda_{opt} \cdot 100$ |
|--------|---------------------|--------------------|-----------------|--------------------------|---------------------------|
| Gc4 | 0.47 | 0.84 | 1.24 | 4.06 | 17.13 |
| PdeL | 0.08 | 0.16 | 0.43 | 1.86 | 5.41 |

Table 3. Average runtimes in seconds of the proposed multilabel segmentation algorithms with respect to differently scaled smoothness values. λ_{opt} is the optimal smoothness parameter with respect to the benchmark.

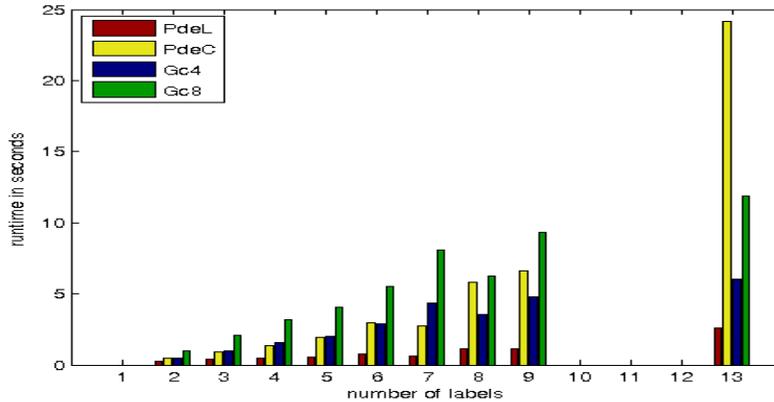


Fig. 4. Comparison of runtimes for different numbers of labels for PDE optimization schemes (PdeL and PdeC) compared to the alpha expansion optimization (Gc4 and Gc8) for the problem of interactive image segmentation.

6 Conclusion

In this paper we proposed an algorithm for interactive multi-region segmentation which takes into account the spatial dimension of the user scribbles. In this way, overlapping color distributions become locally separable allowing for weaker regularization assumptions and correct segmentations in difficult images. We provide an experimental comparison of discrete and continuous optimization approaches. While all algorithms provide similarly good qualitative results on a recently proposed benchmark, the PDE-based methods provide slightly more accuracy, partly due to an absence of metrication errors and partly due to the fact that the multilabel problem is solved by minimizing a single convex energy rather than an iterative sequence of binary problems. Especially the approach by Lellmann et al. is close to real-time on average and runs substantially faster.

References

1. H. Akaike. An approximation to the density function. *Ann. Inst. Statist. Math.*, 6:127–132, 1954.
2. K. Alahari, P. Kohli, and P. Torr. Dynamic hybrid algorithms for map inference in discrete mrfs. *Trans. Pattern Anal. Mach. Intell.*, 32(10):1846–1857, 2010.
3. X. Bai and G. Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. In *Proc. of ICCV*, 2007.
4. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in computer vision. In A. J. M. Figueiredo, J. Zerubia, editor, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 2134 of *LNCS*, pages 359–374. Springer, 2001.

5. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 23(11):1222–1239, 2001.
6. A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Technical report TR-2008-05, University of Bonn, 2008.
7. L. Goldschlager, R. Shaw, and J. Staples. The maximum flow problem is log space complete for p . *Theoretical Computer Science*, 21:105–111, 1982.
8. D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum *a posteriori* estimation for binary images. *J. Roy. Statist. Soc., Ser. B.*, 51(2):271–279, 1989.
9. M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *Proc. of ECCV*, 2008.
10. N. Komodakis and G. Tziritas. A new framework for approximate labeling via graph cuts. In *Proc. of ICCV*, pages 1018–1025, 2005.
11. J. Lellmann, D. Breitenreicher, and C. Schnörr. Fast and exact primal dual iterations for variational problems in computer vision. In *Proc. of ECCV*, 2010.
12. J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multiclass image labeling by simplex-constrained total variation. In *Technical Report, HCI, IWR, University of Heidelberg*, 2008.
13. C. Nieuwenhuis, B. Berkels, M. Rumpf, and D. Cremers. Interactive motion segmentation. In *Proc. of the DAGM*, volume 6376, pages 483–492, 2010.
14. T. Pock, A. Chambolle, D. Cremers, and H. Bischof. A convex relaxation approach for computing minimal partitions. In *Proc. of CVPR*, 2009.
15. R. B. Potts. Some generalized order-disorder transformations. *Proc. Camb. Phil. Soc.*, 48:106–109, 1952.
16. F. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.
17. C. Rother, V. Kolmogorov, and A. Blake. Grab-cut: interactive foreground segmentation using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309–314, 2004.
18. J. Santner. *Interactive Multi-label segmentation*. PhD thesis, University of Graz, 2010.
19. J. Santner, T. Pock, and H. Bischof. Interactive multi-label segmentation. In *Proc. of ACCV*, 2010.
20. R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, A. Agarwala, and C. Rother. A comparative study of energy minimization methods for markov random fields. In *Proc. of ECCV*, pages 16–29, 2006.
21. M. Unger, T. Pock, W. Trobin, D. Cremers, and H. Bischof. Tvseg - interactive total variation based image segmentation. In *Proc. of BMVC*, 2008.
22. 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.