

Spatially Varying Color Distributions for Interactive Multilabel Segmentation

Claudia Nieuwenhuis and Daniel Cremers

Abstract—We propose a method for interactive multilabel segmentation which explicitly takes into account the spatial variation of color distributions. To this end, we estimate a joint distribution over color and spatial location using a generalized Parzen density estimator applied to each user scribble. In this way, we obtain a likelihood for observing certain color values at a spatial coordinate. This likelihood is then incorporated in a Bayesian MAP estimation approach to multiregion segmentation which in turn is optimized using recently developed convex relaxation techniques. These guarantee global optimality for the two-region case (foreground/background) and solutions of bounded optimality for the multiregion case. We show results on the GrabCut benchmark, the recently published Graz benchmark, and on the Berkeley segmentation database which exceed previous approaches such as GrabCut [32], the Random Walker [15], Santner’s approach [35], TV-Seg [39], and interactive graph cuts [4] in accuracy. Our results demonstrate that taking into account the spatial variation of color models leads to drastic improvements for interactive image segmentation.

Index Terms—Image segmentation, spatially varying, color distribution, convex optimization

1 INTRODUCTION

1.1 Interactive Image Segmentation

SEGMENTATION denotes the task of dividing an image into meaningful, nonoverlapping 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 user’s 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 objects in a video, or to interactively edit and manipulate images. Over the last few years, we have observed a number of breakthroughs in image segmentation regarding algorithmic approaches to efficiently compute minimum energy solutions for respective cost functions, using graph cuts [16], [5], level set methods [11], random walks [15], and convex relaxation techniques [8], [42], [19], [7].

Despite substantial algorithmic advances, state-of-the-art approaches to interactive image segmentation often fail for scenes of complex color variability, where objects have similar colors and may be exposed to difficult lighting conditions. The reason is that existing approaches often do not systematically exploit the spatial location of the user information, and rather model the color variation in a given region with a single space-independent color distribution. Due to the strong overlap of respective color distributions,

the segmentation process often fails—see Fig. 2b. In contrast, if only scribble distance is used for region assignment, segmentation fails as well—see Fig. 2c. By statistically taking into account the local distribution of the scribbles we obtain spatially varying color distributions which are locally separable and allow for drastic improvements in the segmentation—see Fig. 2d.

To obtain a good segmentation result most algorithms require two important concepts: First, based on the user input, each pixel is assigned a value measuring how well it fits to each marked region (data fidelity term); second, the consistency of the segmentation with respect to some prior knowledge is imposed (regularization term), e.g., the object boundary length [33], the number of labels [41], specific interclass cost functions [19], or label co-occurrence [20]. In this paper, we focus on the data fidelity term and its importance for the segmentation of natural images taken under difficult color and illumination conditions such as those in Fig. 1. A preliminary version of this work was presented in [25].

1.2 Previous Data Fidelity Terms

Over the years a variety of data fidelity terms for image segmentation have been proposed. The piecewise constant Mumford-Shah functional [23], for example, is based on representing the color in each region with the mean color value, thus minimizing the color variance in each segment. Instead of a constant mean color value, foreground and background histograms are estimated by Boykov and Jolly [4]. Unger et al. [39] also estimate color histograms in their TV-Seg framework and make a hard decision for foreground/background by means of a user threshold on the histogram distance. Mixtures of Gaussians are employed to approximate these probability distributions for digital matting by Chuang et al. [9], in the GrabCut segmentation approach by Rother et al. [32], and for automatic segmentation of natural images by Tai et al. [37]. Methods

• The authors are with the Faculty of Computer Science, Technical University of Munich, Boltzmannstraße 3, 85748 Garching, Germany. E-mail: {claudia.nieuwenhuis, daniel.cremers}@in.tum.de.

Manuscript received 11 Oct. 2011; revised 7 May 2012; accepted 25 July 2012; published online 17 Aug. 2012.

Recommended for acceptance by M. Brown.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2011-10-0731.

Digital Object Identifier no. 10.1109/TPAMI.2012.183.

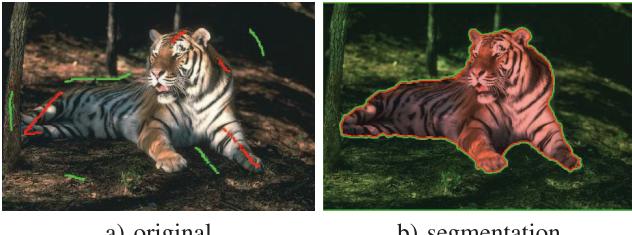


Fig. 1. Interactive segmentation result. (a) The original image exhibits strongly overlapping foreground and background color distributions due to large lighting variations. (b) Explicitly taking into account the spatial dependency of color likelihoods leads to drastic improvements of the segmentation.

from machine learning are also applied to solve the classification task, e.g., support vector machines [12] or random forests [35].

In general, scribbles contain two kinds of information: color and location. Disregarding either one of these two categories leads to suboptimal results, as shown in Fig. 2. With appropriate user interaction the distance to a scribble may convey information on region affiliation. Grady proposes to model this aspect with a random walk [15]. Bai et al. [2] estimate foreground and background color likelihoods whose gradients are used to define weighted geodesic distances between each pixel and the user scribbles. Spatial information has also successfully been introduced into kernel density estimators, e.g., in the mean-shift approach [10], in segmentation based on geodesic distances [2], or in motion analysis [24]. In [40], [30], [14], [17] distance-based sampling was applied to image matting. To estimate the pixel's alpha value, the minimum difference between the linearly interpolated color from each

foreground and background sample and the current pixel's color is sought. In [17], an additional penalty is added for samples at large distances from the current location. In contrast to these approaches, in this paper we estimate a joint probability distribution over color and space from the whole set of samples simultaneously.

In unsupervised segmentation the approaches by Taron et al. [38] and Brox and Cremers [6], possibly the most closely related to ours, are alternating schemes of segmentation and color model estimation. While spatial variation of color distributions often degrades the convergence of such *unsupervised* segmentation methods to suboptimal local minima (the respective color distributions simply adapt to the local context), we show that in *interactive* segmentation—combined with recent convex relaxation techniques—explicitly modeling the spatial context of color information leads to drastic performance improvements with optimal or near-optimal results.

1.3 Statistical Models of Spatial Color Variation

In the interactive segmentation scenario the user scribbles contain information on color and space. By disregarding this spatial information in respective density estimates from user scribbles, purely color-based approaches implicitly assume that the color likelihood is independent of the spatial location. In real-world scenarios this is obviously a restrictive assumption, in particular for natural images with varying light conditions, and for multicolored objects that often rely on camouflaging by adopting similar colors as their environment. Yet, the user scribbles do contain spatial information, albeit in a nontrivial highly correlated form. The aim of this work is to systematically exploit this space-dependency of user scribbles in a Bayesian framework for

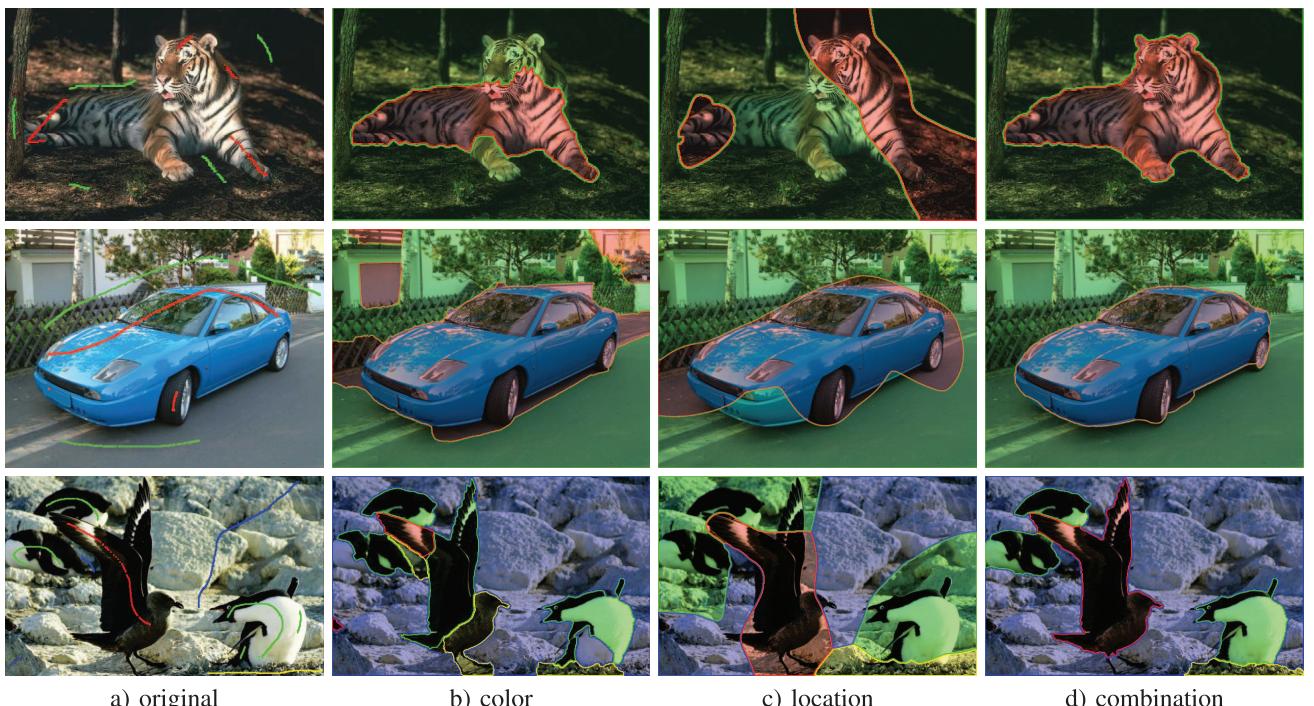


Fig. 2. (a) Original images with strong lighting variations and reflections. (b) Segmentation results based on color similarity only. (c) Segmentation results based on scribble distance only (i.e., a length-regularized Voronoi tessellation). (d) Segmentation results based on spatial and color information of the scribbles.

image segmentation. Experiments confirm that modeling the spatial dependency of color likelihoods comes at no extra cost for the subsequent inference algorithms, preserving the same computation times and optimality guarantees, yet it drastically improves segmentation results.

1.4 Contribution: Bridging the Gap between Distance-Based and Color-Based Segmentation

The contribution of this paper is to demonstrate the importance of regarding not only the color but also the spatial location and distribution of user scribbles in interactive segmentation. The locations of user scribbles are typically highly correlated (rather than uniformly distributed). We present a strategy to account for this violation of the i.i.d. assumption of samples when estimating kernel densities in the joint space of color and location. The resulting approach bridges the gap between purely color-based and purely scribble location-based data terms and thus generalizes previous approaches within one framework. More specifically, we will see later on that our approach has two interesting limiting cases:

- In one limit, we obtain a segmentation method using the commonly employed space-independent color likelihood (namely in disregard of spatial information).
- In another limit, we obtain a length-regularized Voronoi tessellation (in disregard of color information).

A user-defined parameter balances the influence of color and location information. This leads to strongly improved segmentation results within a statistically sound approach.

The proposed framework allows us to segment objects with heavily overlapping color distributions. The approach is formulated as an energy minimization problem which is solved in a variational framework. The solution is provably optimal for two regions, while for multiple regions it lies within small bounds of about less than 1 percent from the optimal solution. The proposed approach can be easily parallelized with computation times of 1.5 seconds on average and is thus well suited for interactivity. Extensive experiments on published benchmarks show that taking into account spatial scribble information outperforms previous approaches, some of which use not only color but also textural information.

2 A STATISTICAL FRAMEWORK FOR SEGMENTATION

In this section, we will derive a Bayesian inference formulation for multiregion segmentation in which space-variant color likelihoods are estimated from a set of user scribbles. One of the challenges we address is that the locations of user scribbles are not i.i.d. distributed but spatially correlated.

2.1 Segmentation as Bayesian Inference

Let $I : \Omega \rightarrow \mathbb{R}^d$ denote the input image defined on the domain $\Omega \subset \mathbb{R}^2$. 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 $u : \Omega \rightarrow \{1, \dots, n\}$, indicating which of the n regions each pixel belongs to: $\Omega_i = \{x | u(x) = i\}$. In the framework of Bayesian inference, one can compute such a segmentation by maximizing the conditional probability:

$$\arg \max_u \mathcal{P}(u | I) = \arg \max_u \mathcal{P}(I | u) \mathcal{P}(u). \quad (2)$$

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

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

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

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

2.2 Inferring Space-Variant Color Distributions

The expression $\mathcal{P}(I(x), x | u(x) = i)$ in (4) denotes the joint 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

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

denote the set of m_i user-labeled pixels x_{ij} and corresponding color values I_{ij} associated with a given region Ω_i . Then, we can estimate the joint distribution $\hat{\mathcal{P}}$ on the product space of color and location by means of a kernel density estimator [1], [31] of the form:

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

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

Fig. 3 shows an example of a distribution in the joint space of color and spatial coordinate estimated from a set of user scribbles. Commonly, the location of the scribbles is not taken into account and the space-independent color distribution, the marginal

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

is used, which is plotted on the right. In this case, we obtain three peaks, one each 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

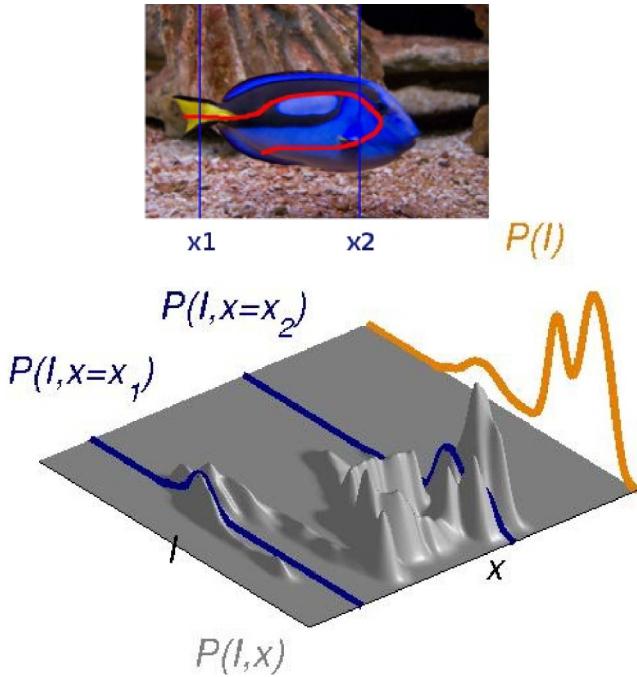


Fig. 3. Top: Fish image with user scribbles indicated in red. Bottom: Scribble-point-based kernel density estimate of the 5D joint distribution of 3D color and 2D location projected to the gray value I and horizontal location x (for readability). If space is not taken into account (i.e., if we marginalize over location), we obtain the same density function $P(I)$ at each pixel, showing several peaks for color values which are predominant in different parts of the object, like yellow, blue, and black. Using the joint distribution gives the probability $P(I, x)$ of observing a color I at a location x . At each location x we obtain a separate color distribution showing only a single color peak for the locally predominant color, e.g., the yellow color in the fin of the fish at horizontal location x_1 or the darker colors like blue and black in the head of the fish at x_2 . The joint 5D distribution over color and space is used as a dataterm for image segmentation. For challenging real-world images it is drastically more informative than the commonly used space-independent distribution $P(I)$.

marginal distribution, no matter if we are very close to a scribble or far away. In contrast, the proposed distribution allows us to distinguish that certain colors may be more or less likely in different locations of the same region Ω_i . Hence, at each location we obtain a separate color distribution, depending on the distance to the scribbles. Clearly, such relevant information on the spatial variation of a color distribution is entirely lost in the traditional space-independent formulations where the spatial dependency is simply integrated out.

The negative logarithm of the estimated probability distribution can be used as a powerful dataterm for segmentation. For each region Ω_i we obtain the dataterm

$$f_i = -\log \hat{P}(I(x), x | u(x) = i). \quad (8)$$

In practice, we choose Gaussian kernels with widths σ and ρ_i in color and space dimension to define the joint probability distribution:

$$\hat{P}(I(x), x | u(x) = i) = \frac{1}{m_i} \sum_{j=1}^{m_i} k_{\rho_i}(x - x_{ij}) k_\sigma(I - I_{ij}). \quad (9)$$

The proposed formulation can be seen as a generalization of the traditional purely color-based approaches. More specifically, it has two interesting limiting cases:

- With larger value of ρ_i the influence of the scribble *locations* becomes less important. In the limit $\rho_i \rightarrow \infty$ we obtain the purely color-based probability:

$$\hat{P}(I(x), x | u(x) = i) = \frac{1}{m_i} \sum k_\sigma(I - I_{ij}). \quad (10)$$

- With increasing values of σ , the influence of scribble *colors* becomes negligible. In the limit $\sigma \rightarrow \infty$ we obtain a purely distance based likelihood:

$$\hat{P}(I(x), x | u(x) = i) = \frac{1}{m_i} \sum k_{\rho_i}(x - x_{ij}). \quad (11)$$

In the subsequent segmentation process this will lead to a regularized Voronoi tessellation.

Hence, by steering ρ_i and σ , the variance of the kernel density estimators, we can scale the influence of color similarity and scribble distance on the segmentation result.

A variant of the proposed approach can be obtained by using the conditional distribution instead of the joint distribution. The conditional distribution for observing a color I given at coordinate x is obtained by normalizing the joint distribution:

$$\hat{P}(I(x) | x, u(x) = i) = \frac{\hat{P}(I(x), x | u(x) = i)}{\int \hat{P}(I(x), x | u(x) = i) dI}. \quad (12)$$

The normalization simply assures that for every point $x \in \Omega$ the probability of observing different colors integrates to 1. Due to the relation

$$P(I, x | u(x) = i) = P(I | x, u(x) = i) P(x | u(x) = i), \quad (13)$$

the joint distribution differs from the conditional by the location prior $P(x | u(x) = i)$ saying if a specific label is likely in this part of the image. Via this prior the pure scribble distance influences the joint probability distribution leading to Voronoi results in the extreme case, as in Fig. 2c. In contrast, in the conditional distribution the normalization neglects this location prior leading to a separate color probability distribution at each pixel. In our experiments, both approaches have shown similar results as, for small values of σ , the location prior does not have large impact on the dataterm. In this paper, we will concentrate on the joint distribution. This generally gives us more freedom in modeling the dataterm since the influence of location and color can be steered separately by means of α and σ . In contrast, with the conditional distribution (i.e., without location prior) the case of a purely distance-based segmentation ($\sigma \rightarrow \infty$) cannot be modeled since we would obtain a uniform distribution.

2.3 Handling Non-iid Samples

Provided that the samples S are independent and identically distributed (iid), the estimator \hat{P} in (9) provably converges to the true distribution for $m_i \rightarrow \infty$ [36]. Unfortunately, this independence assumption is not fulfilled in practice: While observed color values I_{ij} may be assumed to be independent, the spatial coordinates x_{ij} are certainly not. In particular, the samples are not uniformly distributed in space but given by the scribble pattern. To account for this nonuniformity, we employ spatially

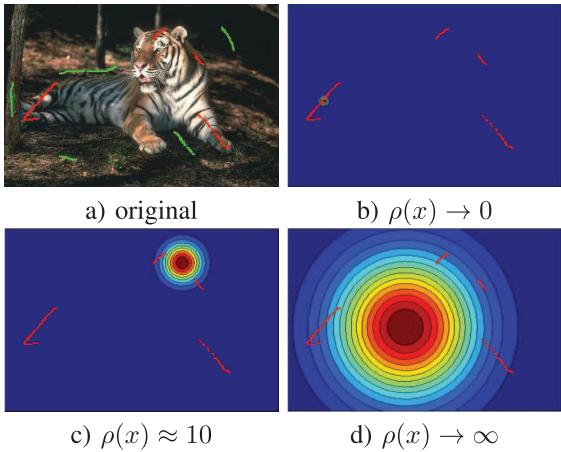


Fig. 4. Influence of scribbles (red) for foreground color density estimation at different image locations. (b) Close to the scribble $\rho(x)$ tends to 0 and only the closest scribble points influence the color distribution. (c) Further away from the scribbles several scribbles in the vicinity influence the estimated color distribution. (d) Far away all scribbles have (almost) equal influence on the estimated color distribution, similar to the case of spatially constant color models.

adaptive kernel functions by choosing the spatial kernel width $\rho_i(x)$ at location x proportional to the distance from the nearest sample point $x_{v_i} \in S_i$ of region i :

$$\rho_i(x) = \alpha|x - x_{v_i}|_2. \quad (14)$$

The effect of this locally dependent spatial variance is demonstrated in Fig. 4. The further away a pixel is from the nearest scribble, the more widespread the region

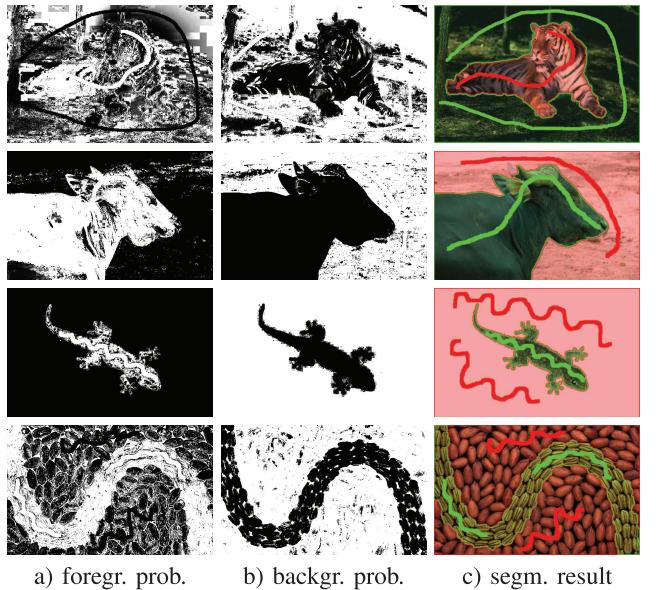


Fig. 6. Label probability $\hat{P}(I(x), x | u(x) = i)$ as defined in (9) for different images from the Berkeley and Graz benchmarks: (a) probability for foreground label, (b) probability for background label, (c) segmentation result

becomes where scribble points are taken into account for color density estimation at this point. In this way, we obtain locally separable space-variant color distributions as shown in Fig. 5.

For several examples, Fig. 5b shows samples randomly drawn from the commonly estimated *spatially constant* color distributions of foreground and background. The images are challenging in the sense that the spatially constant distributions strongly overlap, making it hard or impossible even for humans to detect the object edges correctly. In contrast, Fig. 5c shows randomly drawn samples from *space-variant* color distributions of foreground and background. These reflect the strong variations of color and light in the images. The local adaptation of the probability distributions makes the boundaries of the objects more apparent and ultimately leads to the desired segmentation results, e.g., in Fig. 2d.

Fig. 6 shows the label probabilities as defined in (9) for different images from the Berkeley and Graz benchmarks.

2.4 Variational Formulation

Having determined the probability distributions $\hat{P}(I, x | u(x) = i)$ from the user scribbles for all regions Ω_i , $i = 1, \dots, n$, we are now ready to solve the optimization problem (2). To this end, we specify the prior $\mathcal{P}(u)$ to favor segmentation regions of shorter boundary:

$$\mathcal{P}(u) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^n \text{Per}_g(\Omega_i)\right), \quad (15)$$

where $\text{Per}_g(\Omega_i)$ denotes the perimeter of each set $\Omega_i = \{x \in \Omega | u(x) = i\}$ measured with either an edge dependent or a euclidean metric defined by the nonnegative function $g : \Omega \rightarrow \mathbb{R}^+$:

$$g(x) = \exp(-\gamma|\nabla I(x)|), \quad (16)$$



Fig. 5. Random samples from the estimated foreground and background distributions for the tiger image in Fig. 1a and others. Neglecting the scribble location (b) makes it hard to distinguish between foreground and background and extract the exact boundary line. If the scribble location is regarded (c), the estimated distributions are locally separable and the synthesized image better approximates the original image colors.

for example, favors the coincidence of object and image edges. Maximizing the a posteriori distribution (2) can be solved equivalently by minimizing its negative logarithm. Up to constants and according to (15) and (8), the maximum a posteriori estimation is equivalent to the minimization of the energy

$$\mathcal{E}(\Omega_1, \dots, \Omega_n) = \frac{1}{2} \sum_{i=1}^n \text{Per}_g(\Omega_i) + \lambda \sum_{i=1}^n \int_{\Omega_i} f_i(x) dx, \quad (17)$$

with $f_i(x) = -\log \hat{\mathcal{P}}(I(x) | x | u(x) = i)$.

λ is a weighting parameter which regulates the influence of the dataterm.

3 MINIMIZATION VIA CONVEX RELAXATION

The optimization problem (17) is a classical shape optimization problem, the continuous equivalent of the Potts model [29], which is known to be NP hard. Building upon recent developments in optimization [8], [42], [19], [7], [27], [18], [28], we solve this problem by means of a convex relaxation strategy with the following favorable properties:

- The segmentation is independent of initialization.
- We obtain globally optimal segmentations for the case of two regions and near-optimal—in practice often globally optimal—solutions for the multiregion case.
- The algorithm can be efficiently implemented on the GPU with average computation times of 0.43 seconds, which makes it suitable for realtime user interactions.

3.1 Conversion to a Convex Differentiable Problem

The key idea is to represent the n regions Ω_i by the indicator function $\theta \in \text{BV}(\Omega, \{0, 1\})^n$, where

$$\theta_i(x) = \begin{cases} 1, & \text{if } x \in \Omega_i \\ 0, & \text{otherwise} \end{cases} \quad \forall i = 1, \dots, n. \quad (18)$$

Here, BV denotes the functions of bounded variation, i.e., functions with a finite total variation. In order to rewrite energy (17) in terms of the indicator functions θ_i , let us observe that the boundary of the set indicated by θ_i can be written by means of the total variation. Let $D\theta_i$ denote the distributional derivative of θ_i (which is $D\theta_i = \nabla\theta_i dx$ for differentiable θ_i), $\xi_i \in C_c^1(\Omega, \mathbb{R}^2)$ the dual variables, with C_c^1 the space of smooth functions with compact support.

Then, following the coarea formula [13] the weighted perimeter of all regions Ω_i is equivalent to the weighted total variations:

$$\frac{1}{2} \sum_{i=1}^n \text{Per}_g(\Omega_i) = \frac{1}{2} \sum_{i=1}^n \text{Per}_g(\{x | \theta_i(x) = 1\}) \quad (19)$$

$$= \frac{1}{2} \sum_{i=1}^n \text{TV}_g(\theta_i) \quad (20)$$

$$= \frac{1}{2} \sum_{i=1}^n \int_{\Omega} g |D\theta_i| \quad (21)$$

$$= \sup_{\xi \in \mathcal{K}_g} \sum_{i=1}^n \int_{\Omega} \xi_i D\theta_i \quad (22)$$

$$= \sup_{\xi \in \mathcal{K}_g} \left(- \sum_{i=1}^n \int_{\Omega} \theta_i \operatorname{div} \xi_i dx \right)$$

$$\text{with } \mathcal{K}_g = \left\{ \xi \in C_c^1(\Omega, \mathbb{R}^2)^n \mid |\xi_i(x)| \leq \frac{g(x)}{2}, \right. \\ \left. x \in \Omega, i = 1, \dots, n \right\}. \quad (23)$$

The latter transformation (23) follows from integration by parts and the compact support of the dual variables ξ_i .

For segmentation, we require that at each location $x \in \Omega$ the sum of all indicator functions amount to 1, so each pixel is assigned to exactly one label. With this notation, (17) is equivalent to

$$\min_{\theta \in \mathcal{B}} \mathcal{E}(\theta) = \quad (24)$$

$$\min_{\theta \in \mathcal{B}} \sup_{\xi \in \mathcal{K}_g} \left\{ \sum_{i=1}^n \lambda \int_{\Omega} \theta_i f_i dx - \int_{\Omega} \theta_i \operatorname{div} \xi_i dx \right\} \quad (25)$$

$$\text{with } \mathcal{B} = \left\{ \theta \in \text{BV}(\Omega, \{0, 1\})^n \mid \sum_{i=1}^n \theta_i = 1 \right\}. \quad (26)$$

To obtain a relaxed convex optimization problem which can be minimized globally, we relax the set \mathcal{B} to the convex set

$$\tilde{\mathcal{B}} = \left\{ \theta \in \text{BV}(\Omega, [0, 1])^n \mid \sum_{i=1}^n \theta_i = 1 \right\}. \quad (27)$$

3.2 Numerical Implementation

To solve the relaxed convex optimization problem, we employ a primal dual-algorithm proposed in [28]. Essentially it consists of alternating a projected gradient descent in the primal variables θ_i with projected gradient ascent in the dual variables ξ_i . In addition, it contains an overrelaxation step in the primal variables (giving rise to auxiliary variables $\bar{\theta}_i$):

$$\begin{aligned} \xi^{t+1} &= \Pi_{\mathcal{K}_g} \left(\xi^t + \tau_d \frac{\partial \mathcal{E}}{\partial \xi} \right) = \Pi_{\mathcal{K}_g} (\xi^t + \tau_d \nabla \bar{\theta}^t), \\ \theta^{t+1} &= \Pi_{\tilde{\mathcal{B}}} \left(\theta^t - \tau_p \frac{\partial \mathcal{E}}{\partial \theta} \right) = \Pi_{\tilde{\mathcal{B}}} (\theta^t + \tau_p (\operatorname{div} \xi^{t+1} - \lambda f)), \\ \bar{\theta}^{t+1} &= \theta^{t+1} + (\theta^{t+1} - \theta^t) = 2\theta^{t+1} - \theta^t, \end{aligned} \quad (28)$$

where Π denotes the projections onto the respective convex sets and τ_p and τ_d the primal and dual step sizes. The projection onto \mathcal{K}_g is straightforward by simple clipping; the projection onto the simplex $\tilde{\mathcal{B}}$ is given in [22]. As shown in [28], the algorithm (28) provably converges to a minimizer of the relaxed problem for sufficiently small step sizes τ_p and τ_d which are chosen according to [26].

By allowing the primal variables θ_i to take on intermediate values between 0 and 1, we may end up with nonbinary solutions. By thresholding these to 0 or 1 we can project solutions back to the binary-valued set \mathcal{B} . This projection is known to preserve optimality in case of two regions [8]. In the multiregion case it allows us to compute a

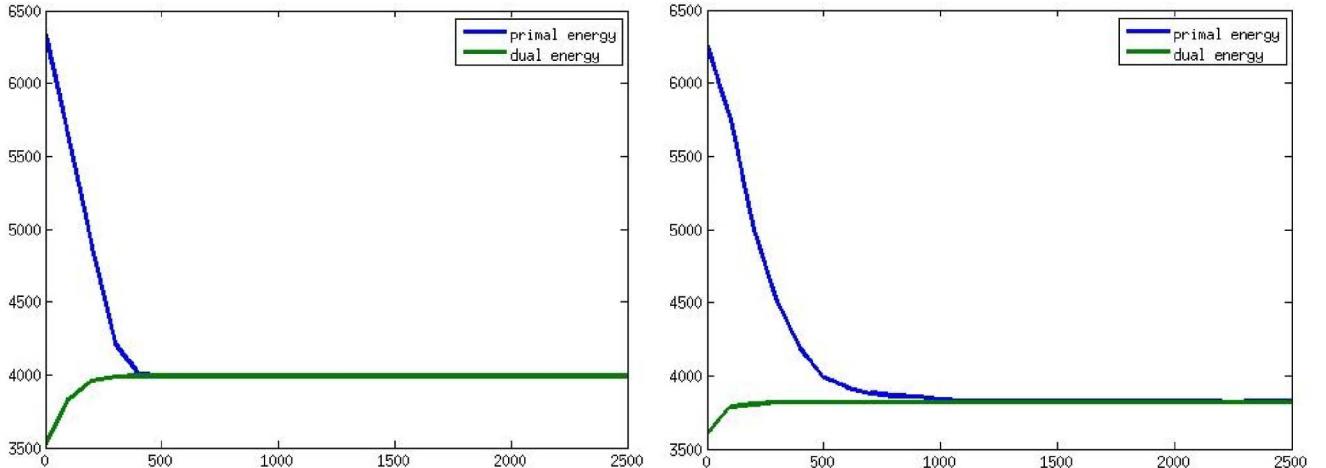


Fig. 7. Convergence analysis of the algorithm for two examples from the Graz benchmark database. The progress of the primal energy (blue) is depicted with respect to the dual energy (green). The optimal solution of the optimization problem is obtained as soon as the primal-dual gap goes to zero.

bound of optimality given by the energy difference between the minimizer of the relaxed problem and its thresholded version (see Proposition 1). Typically, the thresholded solution deviates less than 1 percent from the optimal energy in the multiregion case.

Proposition 1. Let $\theta^* \in \mathcal{B}$ be the global minimizer of the original problem (25), $\tilde{\theta} \in \tilde{\mathcal{B}}$ the result of the proposed algorithm for the relaxed problem, and $\theta' \in \mathcal{B}$ the binarized solution. Then, we can compute an energy bound $\delta(\theta', \tilde{\theta})$ such that the following holds for the energy gap: $E(\theta') - E(\theta^*) \leq \delta(\theta', \tilde{\theta})$.

Proof. Since $\mathcal{B} \subset \tilde{\mathcal{B}}$, we have $E(\tilde{\theta}) \leq E(\theta^*) \leq E(\theta')$. Energetically, the (unknown) optimal binary solution θ^* lies in between the computed relaxed solution $\tilde{\theta}$ and the computed thresholded solution θ' :

$$E(\theta') - E(\theta^*) \leq E(\theta') - E(\tilde{\theta}) =: \delta(\theta', \tilde{\theta}).$$

3.3 Convergence Analysis

To examine the convergence rate, we can compute the primal-dual gap, which is the difference between the primal and the dual energy of the optimization problem. The primal energy is given as the original optimization problem:

$$E_p(\theta) = \left\{ \sum_{i=1}^n \int_{\Omega} \lambda \theta_i f_i dx - g(x) |D \theta_i| \right\} + \delta_{\mathcal{B}}(\theta), \quad (29)$$

where $\delta_{\mathcal{B}}$ is the indicator function of \mathcal{B} :

$$\delta_{\mathcal{B}}(\theta) = \begin{cases} 0, & \theta \in \mathcal{B} \\ \infty, & \theta \notin \mathcal{B} \end{cases} \quad (30)$$

The computation of the dual energy amounts to a pointwise optimization problem:

$$\begin{aligned} E_d(\xi) &= \min_{\theta \in \mathcal{B}} \left\{ \sum_{i=1}^n \int_{\Omega} \theta_i (\lambda f_i - \operatorname{div} \xi_i) dx - \delta_{\mathcal{K}_g}(\xi) \right\} \\ &= \int_{\Omega} \min_i (\lambda f_i - \operatorname{div} \xi_i) dx - \delta_{\mathcal{K}_g}(\xi). \end{aligned} \quad (31)$$

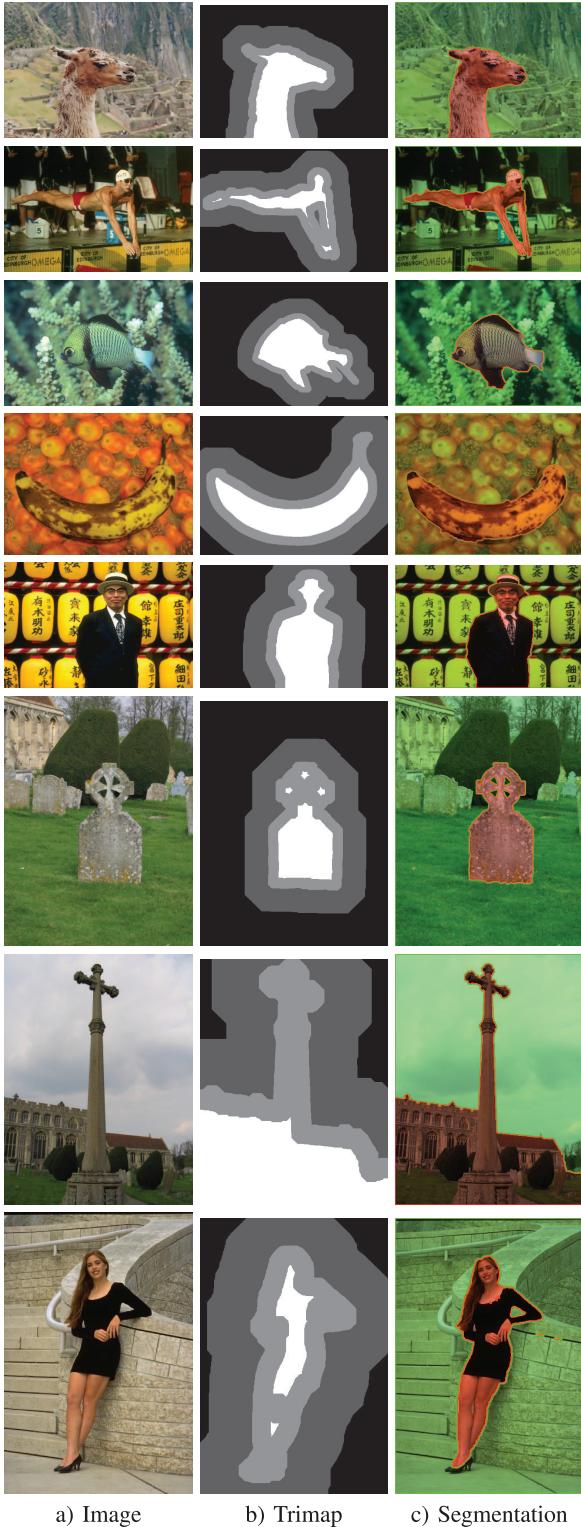
Here, $\delta_{\mathcal{K}_g}$ denotes the indicator function for the set \mathcal{K}_g . During the optimization, the primal energy constantly decreases, whereas the dual energy increases. When the optimal solution is reached, the primal-dual gap goes to zero. The size of the gap can be used to formulate suitable convergence criteria for the algorithm, e.g., the iterations are terminated if the current gap decreases less than 1 percent compared to the previous gap. Fig. 7 shows the primal-dual gap for up to 2,500 iterations for two examples. Alternatively, the change of the solution $|\theta^t - \theta^{t-1}|$ can be measured to determine convergence of the algorithm.

4 RESULTS

We have developed an interactive approach for extracting multiple objects from an image. A key contribution of this paper is to regard both color similarity and the local information of the scribbles. To this end, we introduce space-variant color distributions to model objects with spatially varying appearance in the scene. In the following, we will show experimental results on a variety of real-world images and focus on showing how modeling the spatial variation of color distributions leads to substantial improvements of the segmentation over the traditionally employed spatially constant color distributions. We use the following parameters for space-variant color distributions: $\sigma = 1.3$, $\lambda = 0.008$, $\gamma = 5$, $\alpha = 1.8$, $\tau_p = 0.25$, and $\tau_d = 0.5$ according to [26]. To obtain the spatially constant approach we set α to a very large value, yielding an (approximately) constant spatial distribution.

4.1 Qualitative Benchmark Results

For automatic segmentation, several benchmarks are available, e.g., the Berkeley database, the GrabCut database, or the PascalVocDatabase. As extensively discussed in [34], these benchmarks are not suited for testing interactive segmentation. Hence, Santner et al. [35] published a benchmark for interactive scribble-based segmentation, the Graz benchmark. We nevertheless start our evaluation on the somewhat simpler but widely



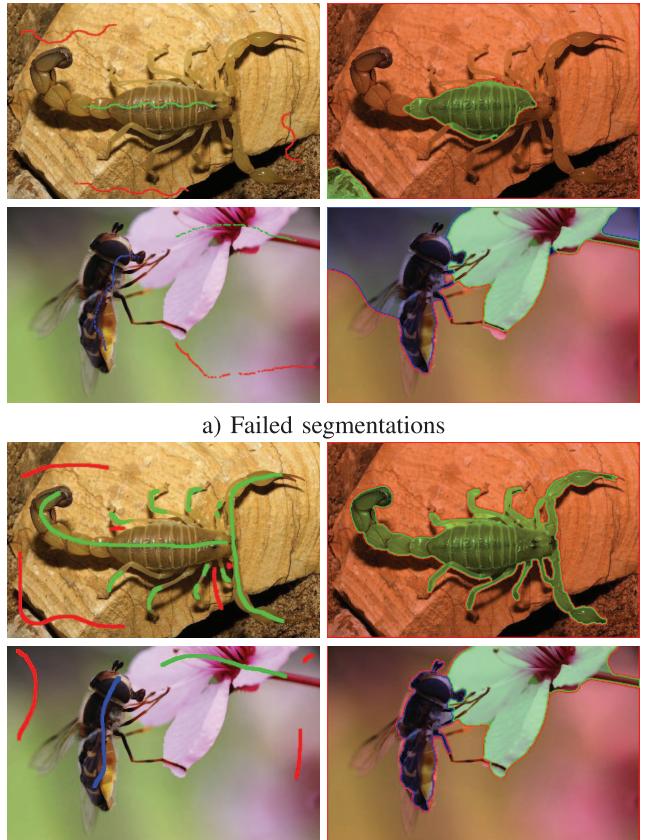
a) Image

b) Trimap

c) Segmentation

Fig. 8. Results on the GrabCut Benchmark. (a) Original image, (b) trimap input (white: foreground, black and dark gray: background, light gray: region to be classified), (c) segmentation result of the proposed approach.

known trimap segmentation problem of the GrabCut benchmark. In a second step we then evaluate the proposed algorithm on the Graz benchmark. We also compare against common segmentation algorithms such as GrabCut and Random Walker.



a) Failed segmentations

b) Improved segmentations

Fig. 9. Failed and improved segmentations. (a) The scribbles are too sparse to adequately reflect the changing colors in the segments. (b) More scribbles have been introduced to obtain the correct segmentations.

Method	E_{miss} in %
GrabCut - Simple Mixture Model [3]	16.3
GrabCut - learned GMMRF [3]	7.9
our approach, space-variant	6.5
Random Walker [15]	1.1

Fig. 10. Comparison to common segmentation approaches such as the Random Walker approach and the GrabCut approach on the standard GrabCut Benchmark based on the misclassification error (in percent) and given trimaps. Segmentation based on trimaps actually represents a simplified problem since the number of objects and the rough boundary

4.1.1 GrabCut Benchmark

For the GrabCut benchmark either bounding boxes enclosing the foreground and—inevitably—part of the background can be used as input or trimaps. Since our approach is based on adaptable color distributions, rough bounding boxes containing both regions are unsuitable for estimating the probability density functions for foreground and background. Hence, we use the trimap input. Trimaps are maps which roughly segment the image except the region around the object boundary, which is to be inferred by the algorithm.



Fig. 11. Results on the Graz benchmark. (a) Approach by Santner et al. [35] based on color and texture, (b) proposed space-variant approach regarding scribble locations.

The benchmark's suitability for testing multilabel segmentation algorithms is limited due to three reasons:

- All examples are limited to two regions, i.e., foreground/background segmentation, which means $n = 2$.
- The objects are assumed to be connected.

- The trimaps contain very large seed regions for foreground and background which already indicate the ground-truth segmentation of most pixels in the image.

Hence, this benchmark simulates strongly simplified conditions for interactive segmentation.

Method	Dim	Brush	Score
Random Walker [15]	3	13	0.855
Santner et al. [35], RGB	3	-	0.877
our approach, spatially constant	3	3	0.889
[35], CIELab + LBP	21	5	0.917
our approach, space-variant	5	5	0.922
[35], CIELab + LBP	21	13	0.927
our approach, space-variant	5	13	0.931

Fig. 12. Comparison to the multilabel segmentation approach by Santner et al. [35] on the Graz benchmark. For each method the dimensionality of the feature vector, the scribble brush radius, and the average dice score are compared. Rather than reverting to sophisticated high-dimensional feature spaces the introduction of spatial information into the estimated color distributions yields better results and allows for higher speeds.

Fig. 8 shows several segmentation results from the Grab-Cut benchmark, the original image on the left, the trimap in the center, and the segmentation result on the right. The segmentation results are very close to the ground truth. Small errors occur due to nonsharp object boundaries (banana), the length regularization (the ear of the llama), or insufficient foreground seed marking (the shoes of the woman).

4.1.2 Graz Benchmark

Since common segmentation databases are not suitable for testing interactive multilabel segmentation [34], Santner et al. recently published the first benchmark for interactive scribble based multilabel segmentation containing 262 seed-ground-truth pairs from 158 natural images containing between two and 13 user labeled segments. Santner et al. [35] show impressive results for different combinations of color and texture features: 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 slightly better than those obtained by Santner et al. (RGB color information without texture). They obtain the best results combining CIELab and LBP features in a 21D vector based on a scribble brush of radius 13. Without reverting to such a sophisticated high-dimensional texture feature space, we obtain a better benchmark performance simply by correctly modeling the spatial variation of color distributions as proposed in this paper.

Fig. 11 shows qualitative benchmark results. The figure contains 16 image pairs, where we contrast Santner et al.'s best results (CIELab colors combined with LBP features) with the results of our space-variant approach. The locality of the color models allows us to release the regularizing model assumptions since we can better distinguish between

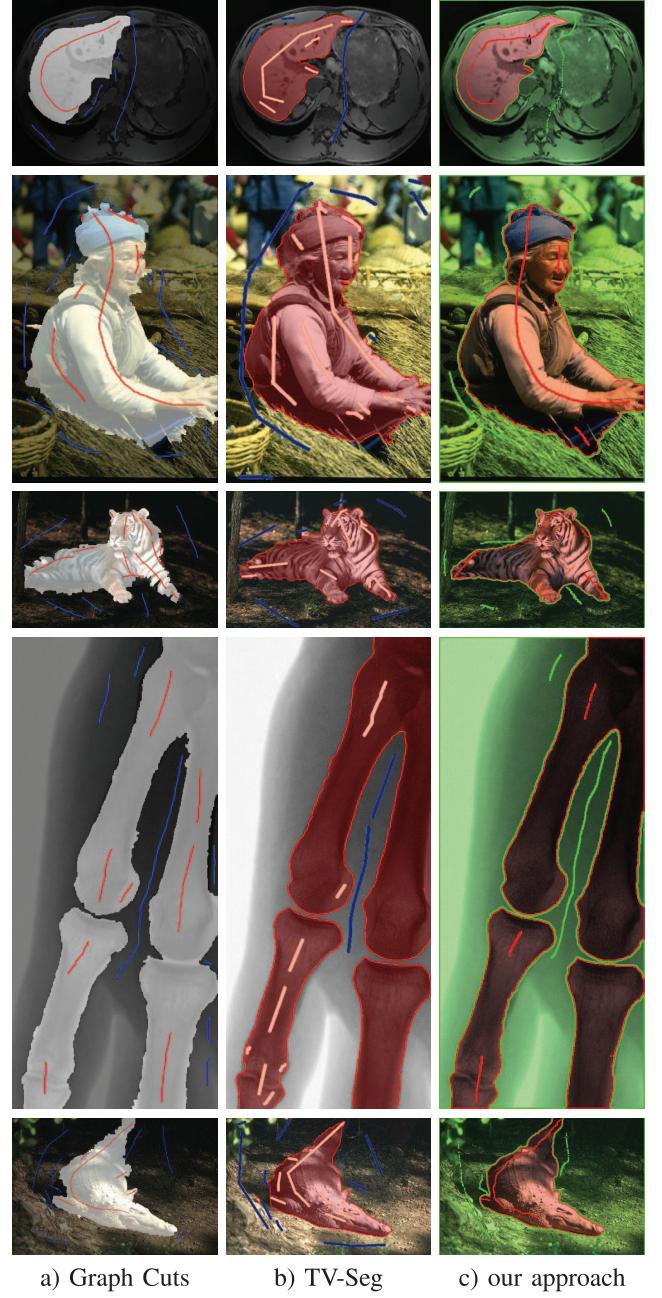


Fig. 12. Comparison to the multilabel segmentation approach by Santner et al. [35] on the Graz benchmark. For each method the dimensionality of the feature vector, the scribble brush radius, and the average dice score are compared. Rather than reverting to sophisticated high-dimensional feature spaces the introduction of spatial information into the estimated color distributions yields better results and allows for higher speeds.

a) Graph Cuts b) TV-Seg c) our approach

Fig. 13. Comparison to known foreground/background extraction approaches. (a) Interactive Graph Cuts [4], (b) TV-Seg [39], (c) the proposed space-variant approach. For (a) and (b) a lot more user interaction is required.

different regions, especially in case of overlapping color models. In this way, for example, the second foot of the duck, the fin of the fish, the head of the man at the lake, the legs of the beacon, the oversmoothed contour of the parrot, the astronaut, and the feathers of the small bird could be correctly recovered. In the case of the snake, the locality of the color model helps to assign the objects to the closer matching segment. This explains the higher accuracy of our approach.

Fig. 9 a shows two examples for failed segmentations which are due to sparse scribbling. The scorpion image would require more scribbles to adequately model the changing color of the animal. In the second image, the red

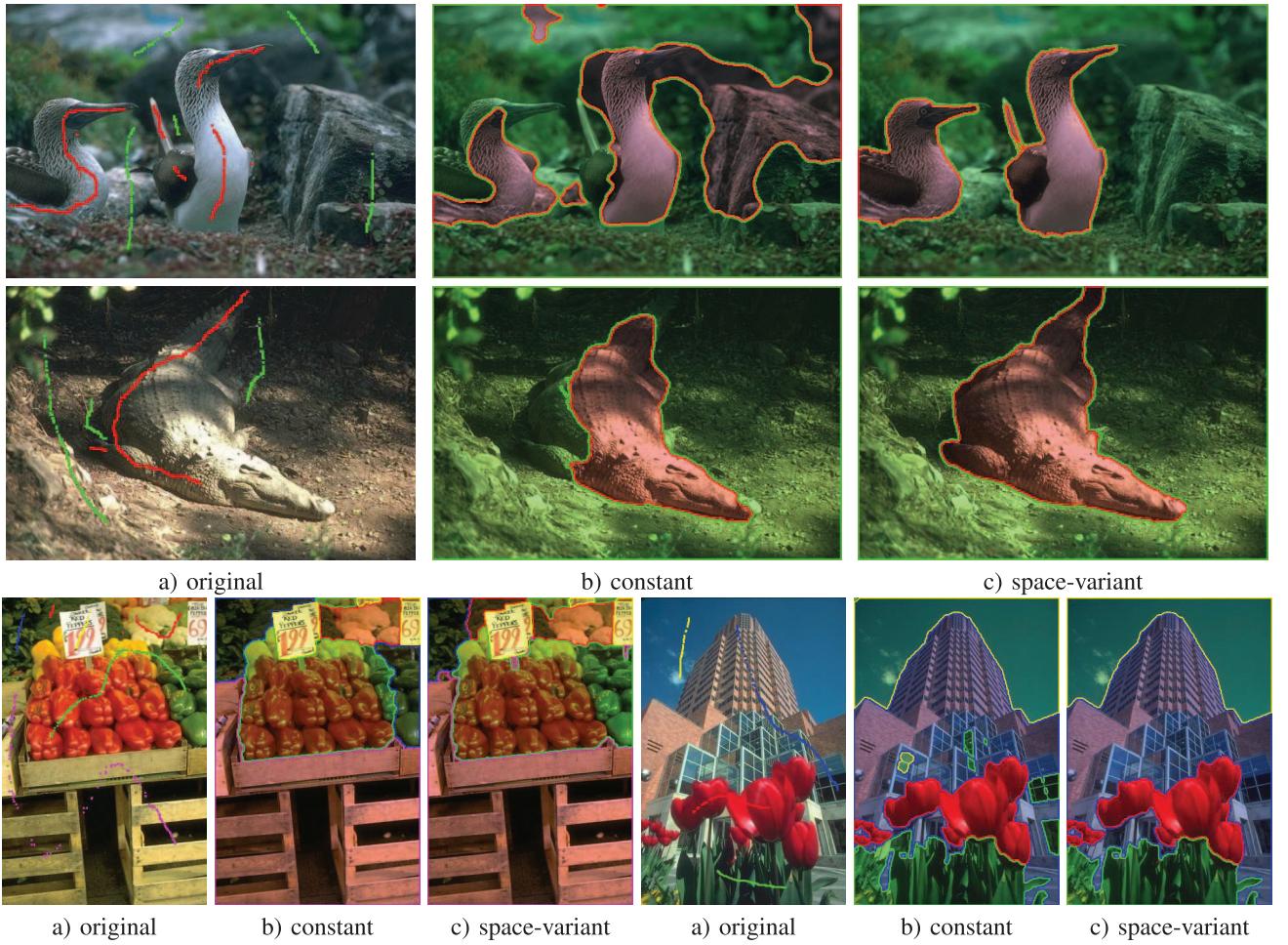


Fig. 14. Multilabel segmentation results for difficult images from the Berkeley and Graz databases. The spatially constant approach (b) is compared to the space variant (c). If spatial scribble information is included in the color model the animals, vegetables, and window reflections can be segmented correctly.

background scribble does not reflect the color change in the upper left image corner. This part of the image is assigned to the foreground due to the strong color similarity with the bee. In Fig. 9b, scribbles have been added to better reflect the color variance of the scorpion and the background of the bee. Based on this additional information, the color distributions are better adapted to the color variation in the images and correct segmentation results can be obtained.

4.2 Quantitative Benchmark Results

We now compare the average quality of the proposed algorithm to previous approaches on the GrabCut and the Graz benchmarks. For each benchmark, different quality measures have been published, which will be explained below.

4.2.1 GrabCut Benchmark

Based on the simplified segmentation problem on the GrabCut benchmark, we obtain the results given in Fig. 10, which indicates the average misclassification error:

$$E_{\text{miss}} = \frac{\text{no. of misclassified pixels}}{\text{no. of pixels in unclassified trimap region}}. \quad (32)$$

From the results we can conclude that the proposed approach outperforms the GrabCut approach [3] by 1.4 percent. However, the Random Walker approach [15] outperforms both algorithms by far. For the Random Walker we chose $\beta = 100$.

4.2.2 Graz Benchmark

To further assess the quality of the proposed algorithm we apply it to the Graz benchmark data set containing 262 images with user scribbles. To evaluate the overall segmentation accuracy Santner et al. [35] compute the arithmetic mean of the Dice-score over all segments. It relates the overlap area of the ground-truth segment $\bar{\Omega}_i$ and the computed segment $\Omega_i = \{x|u(x) = i\}$ to the sum of their areas:

$$\text{dice}(\bar{\Omega}_i, \Omega_i) = \frac{2|\bar{\Omega}_i \cap \Omega_i|}{|\bar{\Omega}_i| + |\Omega_i|}. \quad (33)$$

Fig. 12 compares the average Dice-score of the proposed method to the approach by Santner et al. [35] and the Random Walker [15]. The dimension of the feature vector and the brush size for the scribbles is given. The results indicate that merely regarding the spatial location of scribbles provides stronger performance improvements

than a multitude of sophisticated features. Contrary to the GrabCut benchmark, in the more realistic setting posed by the Graz Benchmark, where only small seeds farther away from the object boundary are given, the Random Walker yields the least accurate results of all algorithms.

4.3 Comparison to Foreground Extraction Tools

Most algorithms for interactive segmentation are limited to two regions, i.e., foreground/background extraction. In this section, we compare our results to those obtained by TV-Seg [39] and Interactive Graph Cuts [4] for images containing only one object in front of the background. Beyond a quantitative analysis of the performance we examined the qualitative performance with varying amounts of scribbles. To demonstrate the effectiveness of space-variant color models we choose images from the Berkeley database with difficult lighting conditions, reflections, or overlapping color models. We used the available implementation of TV-Seg from the TU Graz and the Interactive Graph Cuts implementation in the Toolbox by McGuinness [21]. During segmentation of these images we aimed at minimizing the amount of user scribbles. Fig. 13 shows that all three methods provide segmentations of similar quality. However, the proposed space-variant method better reflects the color variations and thus requires fewer scribbles and user interaction.

4.4 Multilabel Segmentation Results

Finally, we show results for the spatially constant compared to the spatially varying approach for a collection of interesting images from the Berkeley and Graz databases in Fig. 14. The first two rows show different animals in front of a similarly colored background. Segmentation here only succeeds based on color and spatial information. The third row shows the displayed goods of a vegetable stall and a skyscraper. In the vegetable stall, the peppers as well as the cauliflower and broccoli exhibit different colors, but belong to the same segment. Since the color distributions of different segments vary and overlap, it is important to regard the local distribution of the scribbles to obtain a correct segmentation. The right-hand side shows a skyscraper with reflecting windows. Since reflections contain any color from the surroundings, segmentation with global color models is difficult. Based on locally adaptive color models, we obtain correct results.

4.5 Runtimes

Interactive image segmentation algorithms demand runtimes close to real time to be useful for the user. We computed the results on an INTEL XEON 2.27 GHz CPU with an NVIDIA Geforce GTX 580 GPU. The average computation time per image on this benchmark is 1.14 seconds for the dataterm and 0.43 seconds for the optimization for two to 13 labels.

4.6 Sensitivity to Parameters and Scribbles

Finally, we examine the sensitivity of the algorithm with respect to parameter choices and scribble locations.

4.6.1 Parameter Sensitivity

The proposed approach requires the user to select four input parameters: the color kernel variance σ , the distance factor α determining the spatial scribble influence, the

σ	λ	γ	α	Brush	Score
1	0.01	5	0.7	3	0.909
1.5	0.01	5	0.7	3	0.909
1.5	0.01	5	0.7	5	0.910
1.5	0.01	5	1	5	0.917
1.5	0.01	4	1	5	0.917
1.5	0.01	5	1.8	5	0.923
1.5	0.01	5	1.8	13	0.927
1.3	0.005	5	1.8	13	0.912
1.3	0.008	5	1.8	13	0.931

Fig. 15. Sensitivity to parameter modifications. The table shows the average Dice-score on the Graz Benchmark for different sets of the algorithm's parameters. For reasonable choices the performance only varies within 2 percent.

smoothness factor λ regulating the strength of the dataterm, and the parameter γ of the edge indicator function g .

As mentioned above, all experiments reported so far were carried out with a fixed set of parameters. Nevertheless, in order to assess the sensitivity of our algorithm to variations in these parameters, we reran the Graz benchmark performance for a variety of parameter settings. Fig. 15 shows that for variations of these parameters within a reasonable range, the benchmark performance only varies within 2 percent. Among the parameters to be set, the spatial influence of the scribbles α , the dataterm weight λ , and the brush size seem to have the largest impact on the results.

4.6.2 Sensitivity to Scribble Location

Fig. 16 shows segmentation results for different amounts and locations of user scribbles. The segmentation based on the first four input sets is almost identical. The random distribution of scribble points in the last image yields slightly better results for the legs of the animal and the stem of the flower due to the wide-spread sample distribution in the image. We can conclude that provided the scribbles are set reasonably to represent the color distributions of the separate regions, the segmentation quality is not very sensitive to their amount or location.

5 CONCLUSION

In this paper, we proposed an algorithm for interactive multiregion segmentation which takes into account not only color similarity but also the spatial location of the user scribbles. Since correct density estimation requires iid assumptions on the spatial location of the scribbles, which is not the case, we proposed a way to handle this problem. The result is a scalable dataterm which bridges the gap between purely color-based and purely distance-based segmentations. In this way, overlapping color distributions become locally separable, allowing for

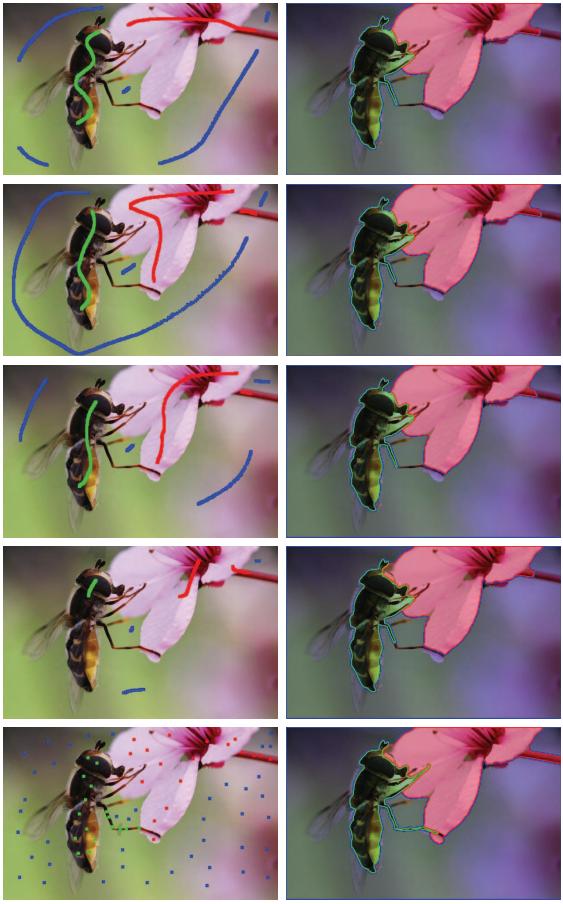


Fig. 16. Segmentation results for different scribble amounts and locations. The results show that the algorithm is not very sensitive to the user input provided the scribbles are placed in reasonable locations to well represent the color variation in the image.

weaker regularization assumptions and correct segmentations in difficult images. The variational approach is derived rigorously in a framework of Bayesian inference and minimized by efficient convex relaxation techniques, leading to globally optimal solutions for the two-region case and solutions of bounded optimality (and, in practice, often optimal solutions) for the multiregion case. The approach can be parallelized, leading to computation times around 1.1 seconds on average for 2 to 13 labels. Results on the GrabCut, Graz, and Berkeley databases show that we exceed previous approaches such as GrabCut, Random Walker, and TV-Seg.

ACKNOWLEDGMENTS

The authors thank Andrew Delong for fruitful discussions of the relations between color-based and space-based distances.

REFERENCES

- [1] H. Akaike, "An Approximation to the Density Function," *Ann. Inst. Statistical Math.*, vol. 6, pp. 127-132, 1954.
- [2] X. Bai and G. Sapiro, "A Geodesic Framework for Fast Interactive Image and Video Segmentation and Matting," *Proc. 11th IEEE Int'l Conf. Computer Vision*, 2007.
- [3] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Interactive Image Segmentation Using an Adaptive GMRF Model," *Proc. European Conf. Computer Vision*, 2004.
- [4] Y. Boykov and M. Jolly, "Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images," *Proc. Eighth IEEE Int'l Conf. Computer Vision*, 2001.
- [5] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222-1239, Nov. 2001.
- [6] T. Brox and D. Cremers, "On Local Region Models and a Statistical Interpretation of the Piecewise Smooth Mumford-Shah Functional," *Int'l J. Computer Vision*, vol. 84, pp. 184-193, 2009.
- [7] A. Chambolle, D. Cremers, and T. Pock, "A Convex Approach for Computing Minimal Partitions," Technical Report TR-2008-05, Dept. of Computer Science, Univ. of Bonn, Germany, 2008.
- [8] T. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models," *SIAM J. Applied Math.*, vol. 66, no. 5, pp. 1632-1648, 2006.
- [9] Y.-Y. Chuang, B. Curless, D.H. Salesin, and R. Szeliski, "A Bayesian Approach to Digital Matting," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 264-271, 2001.
- [10] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach to Feature Space Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603-619, May 2002.
- [11] D. Cremers, M. Rousson, and R. Deriche, "A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape," *Int'l J. Computer Vision*, vol. 72, no. 2, pp. 195-215, Apr. 2007.
- [12] O. Duchenne and J.-Y. Audibert, "Fast Interactive Segmentation Using Color and Textural Information," technical report, Certis, Paris Tech., 2006.
- [13] H. Federer, *Geometric Measure Theory*. Springer, 1996.
- [14] E.S.L. Gastal and M.M. Oliveira, "Shared Sampling for Real-Time Alpha Matting," *Eurographics Computer Graphics Forum*, vol. 29, no. 2, pp. 575-584, 2010.
- [15] L. Grady, "Random Walks for Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768-1783, Nov. 2006.
- [16] D.M. Greig, B.T. Porteous, and A.H. Seheult, "Exact Maximum A Posteriori Estimation for Binary Images," *J. Royal Statistical Soc. Ser. B*, vol. 51, no. 2, pp. 271-279, 1989.
- [17] K. He, C. Rhemann, C. Rother, X. Tang, and J. Sun, "A Global Sampling Method for Alpha Matting," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2049-2056, 2011.
- [18] J. Lellmann, F. Becker, and C. Schnörr, "Convex Optimization for Multi-Class Image Labeling with a Novel Family of Total Variation Based Regularizers," *Proc. 12th IEEE Int'l Conf. Computer Vision*, 2009.
- [19] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr, "Convex Multiclass Image Labeling by Simplex-Constrained Total Variation," technical report, HCI, IWR, Univ. of Heidelberg, 2008.
- [20] P. Kohli, L. Ladicky, C. Russell, and P. Torr, "Graph Cut Based Inference with Co-Occurrence Statistics," *Proc. 11th European Conf. Computer Vision*, 2010.
- [21] K. McGuinness and N. O'Connor, "A Comparative Evaluation of Interactive Segmentation Algorithms," *Pattern Recognition*, vol. 43, no. 1, pp. 434-444, 2010.
- [22] C. Michelot, "A Finite Algorithm for Finding the Projection of a Point Onto the Canonical Simplex of r^n ," *J. Optimization Theory and Applications*, vol. 50, no. 1, pp. 195-200, 1986.
- [23] D. Mumford and J. Shah, "Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems," *Comm. Pure Applied Math.*, vol. 42, pp. 577-685, 1989.
- [24] C. Nieuwenhuis, B. Berkels, M. Rumpf, and D. Cremers, "Interactive Motion Segmentation," *Proc. 32nd DAGM Conf. Pattern Recognition*, vol. 6376, pp. 483-492, 2010.
- [25] C. Nieuwenhuis, E. Töpke, and D. Cremers, "Space-Varying Color Distributions for Interactive Multiregion Segmentation: Discrete versus Continuous Approaches," *Proc. Eighth Int'l Conf. Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2011.
- [26] T. Pock and A. Chambolle, "Diagonal Preconditioning for First Order Primal-Dual Algorithms in Convex Optimization," *Proc. 13th IEEE Int'l Conf. Computer Vision*, 2011.
- [27] T. Pock, A. Chambolle, H. Bischof, and D. Cremers, "A Convex Relaxation Approach for Computing Minimal Partitions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009.

- [28] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, "An Algorithm for Minimizing the Piecewise Smooth Mumford-Shah Functional," *Proc. 12th IEEE Int'l Conf. Computer Vision*, 2009.
- [29] R.B. Potts, "Some Generalized Order-Disorder Transformations," *Proc. Cambridge Philosophical Soc.*, vo. 48, pp. 106-109, 1952.
- [30] C. Rhemann, C. Rother, and M. Gelautz, "Improving Color Modeling for Alpha Matting," *Proc. British Machine Vision Conf.*, 2008.
- [31] F. Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Math. Statistics*, vol. 27, pp. 832-837, 1956.
- [32] C. Rother, V. Kolmogorov, and A. Blake, "Grab-Cut: Interactive Foreground Segmentation Using Iterated Graph Cuts," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 309-314, 2004.
- [33] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D*, vol. 60, pp. 259-268, 1992.
- [34] J. Santner, "Interactive Multi-Label Segmentation," PhD thesis, Univ. of Graz, 2010.
- [35] J. Santner, T. Pock, and H. Bischof, "Interactive Multi-Label Segmentation," *Proc. Asian Conf. Computer Vision*, 2010.
- [36] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1992.
- [37] Y. Tai, J. Jia, and C. Tang, "Soft Color Segmentation and Its Applications," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1520-1537, Sept. 2007.
- [38] M. Taron, N. Paragios, and M.-P. Jolly, "Border Detection on Short Axis Echocardiographic Views Using an Ellipse Driven Region-Based Framework," *Medical Image Computing and Computer-Assisted Intervention*, vol. 3216, pp. 443-450, 2004.
- [39] M. Unger, T. Pock, W. Trobin, D. Cremers, and H. Bischof, "TVSeg—Interactive Total Variation Based Image Segmentation," *Proc. British Machine Vision Conf.*, 2008.
- [40] J. Wang and M. Cohen, "Optimized Color Sampling for Robust Matting," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 2007.
- [41] J. Yuan and Y. Boykov, "TV-Based Image Segmentation with Label Cost Prior," *Proc. British Machine Vision Conf.*, 2010.
- [42] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer, "Fast Global Labeling for Real-Time Stereo Using Multiple Plane Sweeps," *Proc. Vision, Modeling and Visualization Workshop*, 2008.



Claudia Nieuwenhuis received the MS (Diplom) degree in computer science from the Technical University of Ilmenau, Germany, in 2006 and the MS (Diplom) degree in mathematics from the University of Hagen, Germany, in 2008. In 2009, she received the PhD degree in computer science from the University of Heidelberg, Germany. Between 2003 and 2006 she worked for a year and a half at Siemens Corporate Research in Princeton, New Jersey, and at the Fraunhofer Center for Advanced Media Technology in Singapore. After receiving the PhD degree she stayed on as a postdoctoral researcher with Professor Rudolf Mester at the University of Frankfurt. Since October 2009, she has been with the Computer Vision and Pattern Recognition Group at the Technical University of Munich headed by Professor Daniel Cremers.



Daniel Cremers received the MS (Diplom) degree in theoretical physics in 1997 from the University of Heidelberg, Germany, and the PhD degree in computer science in 2002 from the University of Mannheim, Germany. Subsequently, he spent two years as a postdoctoral researcher at the University of California Los Angeles (UCLA) and one year as a permanent researcher at Siemens Corporate Research in Princeton, New Jersey. From 2005 to 2009, he headed the Computer Vision Group at the University of Bonn, Germany. Since 2009, he has been a full professor of computer science and mathematics at the Technische Universität München, Germany. He has received several awards, including the Best Paper of the Year 2003 by the Pattern Recognition Society, the Olympus Award 2004, and the 2005 UCLA Chancellor's Award.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.