TUM

TECHNISCHE UNIVERSITÄT MÜNCHEN
INSTITUT FÜR INFORMATIK

Label Configuration Priors for Continuous Multi-Label Optimization

Mohamed Souiai, Evgeny Strekalovskiy, Claudia Nieuwenhuis, Daniel Cremers

TUM-I1221

Label Configuration Priors
for Continuous Multi-Label Optimization

Mohamed Souiai
Department of Computer Science,
TU Munich
mohamed.souiai@tum.de

Evgeny Strekalovskiy
Department of Computer Science,
TU Munich
evgeny.strekalovskiy@in.tum.de

Claudia Nieuwenhuis
Department of Computer Science,
TU Munich
claudia.nieuwenhuis@in.tum.de

Daniel Cremers
Department of Computer Science,
TU Munich
cremers@tum.de

Abstract

We propose a general class of label configuration priors for continuous multi-label optimization problems. In contrast to MRF-based approaches, the proposed framework unifies label configuration energies such as minimum description length priors, co-occurrence priors and hierarchical label cost priors. Moreover, it does not require any preprocessing in terms of super-pixel estimation. All problems are solved using efficient primal-dual algorithms which scale better with the number of labels than the alpha-expansion method commonly used in the MRF setting. Experimental results confirm that label configuration priors lead to drastic improvements in segmentation. In particular, the hierarchical prior allows to jointly compute a semantic segmentation and a scene classification.

1. Introduction

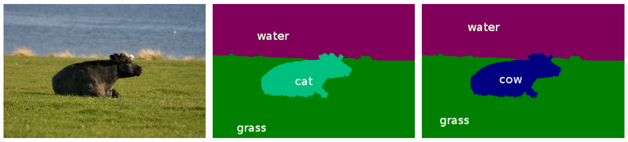
1.1. Semantic Image Labeling

While traditional image segmentation algorithms have focused on separating regions based on homogeneity of color or texture, more recent methods have aimed at incorporating *semantic* knowledge into what is often called class-based image segmentation. Rather than simply grouping regions of similar color, the goal is to assign to each pixel of an image a semantic label such as “grass”, “sky”, “cow” or “horse”, each of which does not necessarily share the same color model – horses may be white, brown or black for example. Such approaches allow to impose prior knowledge about which pairs of labels are likely to co-occur in a given image [8]. Figure 1 shows semantic labelings computed for the image of a cow on grass without and with a co-occurrence prior: While the data term has a slight preference for cat over cow, the co-occurrence additionally imposes the information that cows are more commonly observed on grass next to the ocean than cats.

A separate line of work has promoted the use of minimum description length (MDL) priors [9, 17, 4, 15] which imposes a prior that favors a smaller number of labels in the final segmentation. In practice, the advantage of such MDL priors is that one can preserve a level of regularity while reducing the often oversmoothing boundary length regularization. While many experiments demonstrating the advantage of co-occurrence can often be reproduced with a simple MDL prior (that suppresses the emergence of undesired labels), for the example in Figure 1 co-occurrence is vital (since the number of labels is in both cases the same).

Additionally, hierarchical priors have been introduced in [3]. These favor specific label groups which usually occur in the same context, e.g. a dishwasher and an oven are likely to occur together within the context of a kitchen, whereas a cow and a sheep usually appear in natural contexts outside. Such hierarchical priors are more general than co-occurrence priors since they do not distinguish between the labels within a specific context but only between labels of different contexts thus building a kind of label hierarchy.

A major challenge addressed in this paper is how to efficiently integrate such label configuration priors in a convex continuous optimization approach, which allows for fast solutions which are independent of the initialization of the algorithm.



Input image no prior configuration prior

Figure 1. Whereas purely data-driven semantic segmentation (middle) assigns the label 'cat' to the cow, label configuration priors such as co-occurrence priors (right) can substantially improve the semantic image labeling. The co-occurrence prior imposes the knowledge that cows are more commonly encountered next to grass and ocean than cats.

1.2. Related work

The inspiration to this work predominantly draws from two lines of research, namely research on label configuration priors and research on convex relaxation techniques:

On then one hand, there are a number of recent advances on label configuration energies for semantic image labeling, including the co-occurrence priors by Ladicky et al. [8], the MDL priors of Delong et al. [4] and Yuan et al. [15], and the hierarchical label cost priors proposed by Delong et al. [3]. While the approach by Ladicky et al. demonstrates excellent labeling performance using co-occurrence statistics, it requires a preprocessing in terms of super-pixel estimation. Clearly, this preprocessing step is suboptimal in the sense that the super-pixel estimation has no knowledge of the co-occurrence prior: pixels erroneously combined in a super-pixel will not get separated in the subsequent co-occurrence based optimization process. This effect is shown in Figure 2.

On the other hand, there are a number of recent advances on convex relaxation techniques for spatially continuous multi-label optimization. These include relaxations for the continuous Potts model [1, 10, 16], for the non-local continuous Potts model [14], for MDL priors [15], and for vector-valued labeling problems [6, 13].

1.3. Contribution

Apart from the MDL prior, all label configuration priors have been applied to the spatially discrete MRF domain, which exhibit grid bias and are hard or impossible to parallelize. As discussed in [7], spatially continuous approaches do not exhibit any grid bias and respective algorithms are easily parallelized on graphics hardware.

Hence, we propose a convex framework for multi-label optimization which can incorporate all of the above label configuration priors. To this end, we formulate a variational approach which can be optimized with fast primal-dual schemes. In contrast to the co-occurrence approach by Ladicky et al. [8], the proposed method does not require any pre-processing in terms of super-pixel estimation but directly works on the input pixels. In contrast to the hierarchical prior approach by Delong et al. [3], it scales well with the number of labels – examples with 256 labels are easily handled. Additionally we propose a generalisation of the minimal description length prior given by Yuan et al. [15] to a general formulation by a composition with a with an arbitrary convex and monotonously increasing function.

2. Formulating a Convex Optimization Problem

Given a discrete labelspace $\mathcal{G} = \{1,..,n\}$ with a $n \geq 1$, the multi-labeling problem can be stated as a minimal partition problem. The image domain $\Omega \subset \mathbb{R}^2$ is to be segmented into n of pairwise disjoint regions Ω_i by minimizing a general energy E which can be decomposed as follows:

$$E = E_D + E_S + E_L, \tag{1}$$

The term E_D is called the data term, the expression E_S represents a smoothness regularizer and the term E_L is a global energy which penalizes specific label configurations in the image and will be introduced in Section 3 .

To obtain a convex energy, the label assignment function is expressed in terms of the n label indicator functions $u_i : \Omega \to \{0,1\},\ i \in \{1,..,n\}$ defined by

$$u_i(x) = \begin{cases} 1 & \text{if } x \in \Omega_i, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

To ensure that each pixel is assigned to exactly one region the simplex constraint is imposed on u

$$\sum_{i=1}^{n} u_i(x) = 1 \quad \forall x \in \Omega. \tag{3}$$

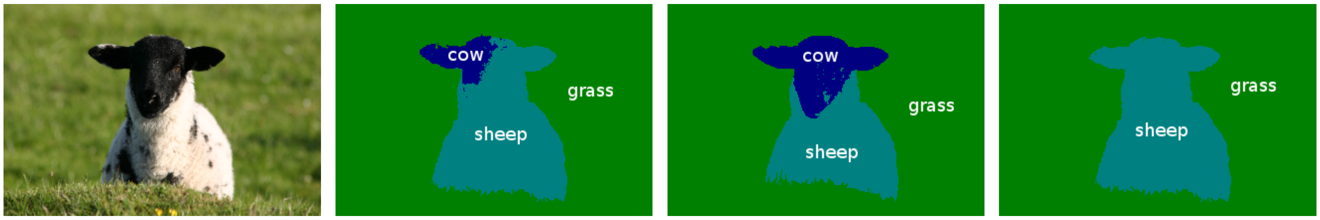
Data Term $E_D(u)$. The data term assigns the cost $\varrho_i(x)$ to each pixel x for belonging to region i, e.g. based on color distance and can be written in terms of the indicator functions as

$$E_D(u) = \sum_{i=1}^{n} \int_\Omega u_i(x)\varrho_i(x)dx. \tag{4}$$

Here, $\varrho_i : \Omega \to \mathbb{R}$ indicates the cost of assigning the label i to pixel x. It can be computed from color models, which can be either indicated by the user or learned from training data.

Smoothness Term $E_S(u)$. The smoothness term imposes a spatial prior which can be formulated by choosing the Potts model:

$$E_S(u) = \frac{1}{2} \sum_{i=1}^{n} \int_\Omega |Du_i(x)| \tag{5}$$



Input no co-occ. Ladicky et al. proposed

Figure 2. The CRF approach by Ladicky et al. [8] requires a super-pixel preprocessing which assigns to the entire head a single super-pixel labeled as "cow". The subsequent co-occurrence based optimization cannot recover from this mistake due to the large number of mislabeled pixels. Since the proposed method imposes co-occurrency directly in the labeling of the input pixels it does not suffer from this drawback.

which penalizes the total interface length. Since u_i is not differentiable in general, we consider its dual formulation

$$E_S(u) = \sup_{p \in K} \sum_{i=1}^{n} \int_{\Omega} u_i(x) \operatorname{div} p_i(x) dx \tag{6}$$

with the convex set

$$K = \left\{ p : \Omega \to (\mathbb{R}^2)^n \mid |p_i(x)| \leq \frac{1}{2} \right\}. \tag{7}$$

The classical TV formulation of the Potts model minimizes the length of the interface of each region which can be an inappropriate prior in images exhibiting fine or elongated structures. Hence, Werlberger et al. [14] proposed a non-local variant of the Potts models, which improves the labeling quality on the boundaries:

$$E_S(u) = \sum_{i=1}^{n} \int_{\Omega} \int_{\mathcal{N}_x} w(x,y) |u_i(y) - u_i(x)| dy\, dx \tag{8}$$

where $\mathcal{N}_x$ indicates a predefined neighborhood of x and $w(x,y)$ defines the support weight between the pixels. It is specified as

$$w(x,y) = \exp\left[-\left(\frac{d_c(x,y)}{\alpha} + \frac{d_p(x,y)}{\beta} \right) \right]. \tag{9}$$

In this way, the color distance $d_c(x,y)$ and the Euclidean distance $d_p(x,y)$ between pixels x and y are combined.

The dual formulation of (8) is given by

$$E_S(u) = \sup_{p \in \tilde{\mathcal{K}}} \sum_{i=1}^{n} \int_{\Omega} \int_{\mathcal{N}_x} p(x,y)(u_i(y) - u_i(x)) dy\, dx \tag{10}$$

with the convex constraint set $\tilde{\mathcal{K}}$

$$\tilde{\mathcal{K}} := \left\{ p(x,y) : \Omega \times \Omega \to \mathbb{R} \,\middle|\, |p(x,y)| \leq w(x,y) \right\}. \tag{11}$$

Thus E_D and E_S above are both convex in u.

Convex Relaxation. In order to obtain a convex optimization problem, the optimization domain must be convex as well. Therefore, we relax the binary constraints $u_i(x) \in \{0,1\}$ together with (3) and obtain the convex set

$$\mathcal{S} := \left\{ (u : \Omega \to \mathbb{R}^n \,\middle|\, \sum_{i=1}^{n} u_i(x) = 1, \;\; u_i(x) \in [0,1] \right\}. \tag{12}$$

3. Label Configuration Priors

In this section we introduce the global energy E_L in (1), which depends only on the presence of a specific label in the image, not on the location or the size of the corresponding region in the segmentation, i.e. label configurations like the number of non-empty regions (MDL prior), the co-occurrences of certain labels (co-occurrence prior) or categories/subgroups of labels (hierarchical prior) occurring in the segmentation are penalized.

To make this precise, we introduce the indicator functions $l = (l_1, .., l_n) : \mathcal{S} \to \{0,1\}^n$,

$$l_i(u) = \begin{cases} 1, & \text{if } \exists x \in \Omega : \; u_i(x) = 1, \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

which indicate if a label i appears in the segmentation given by the indicator functions u.

Our approach can handle general label configuration energies E_L subject to the following properties:

1. $E_L(l)$ is convex.
2. $E_L(l)$ is strictly monotonously increasing in l.

Convex Formulation of E_L To obtain a convex formulation of the label indicator functions l_i, we use the following relation:

$$l_i(u) = \max_{x \in \Omega} u_i(x) \qquad \forall i \in \mathcal{G}. \tag{14}$$

Since the max operator is not differentiable we relax the above equation to $l : \mathcal{S} \to [0,1]^n$ with the convex constraint:

$$l_i(u) \geq u_i(x) \quad \forall x \in \Omega \quad \forall i \in \mathcal{G}. \tag{15}$$

We prove the following theorem, which states that the constraint in (14) can be replaced by the constraint in (15) preserving the optimum of the optimization problem.

Theorem 1. *Let* (u^*, l^*) *be a minimizer of*

$$E(u, l) = E_D(u) + E_S(u) + E_L(l) \tag{16}$$

subject to the convex constraint (15)*. Then* (u^*, l^*) *is also a minimizer of the same energy* (16) *and the constraint* (14) *is recovered.*

Proof. Let (u^*, l^*) be given and satisfying (15). If (u^*, l^*) also satisfies (14) we are done. Otherwise there exists $i \in \{1, .., n\}$ such that $l_i^*(u^*) > \max_{x\in\Omega} u_i^*(x)$. Let $\hat{l}$ be a vector with $\hat{l}_i(u) := \max_{x\in\Omega} u_i^*(x)$. Then $\hat{l}_i < l_i^*$ and

$$\begin{aligned} E(u^*, \hat{l}) &= E_D(u^*) + E_S(u^*) + E_L(\hat{l}) \\ &< E_D(u^*) + E_S(u^*) + E_L(l^*) = E(u^*, l^*), \end{aligned} \tag{17}$$

since E_L is strictly monotonously increasing. Thus, $(u^*, \hat{l})$ has a lower energy than $(u^*, \hat{l})$, contradicting the assumption. □

Next, we will give some examples and implementation details for the label interactions E_L arising frequently in practice.

3.1. A Generalized MDL Prior

The MDL prior can be written as the following label configuration energy:

$$E_L(l) = \sum_{i=1}^{n} l_i(u) C_i, \tag{18}$$

where C_i represents a predefined cost for each label occurring in the image. This prior has been previously considered in [15, 4]. By choosing C_i differently for each $1 \leq i \leq n$ one can penalize subsets of labels differently than others.

We generalize this prior to the following energy

$$E_L(l) = f\left(\sum_{i=1}^{n} l_i(u)\right) \tag{19}$$

with an arbitrary convex and monotonously increasing function $f : \mathbb{R} \to \mathbb{R}$. Specific choices are $f(s) = s$ for the number of labels as in [15, 4], $f(s) = s^2$, or $f(s) = \delta_{s \leq s_0}$ for some $s_0 \geq 1$, i.e. $f(s) = 0$ if $s \leq s_0$ and $f(s) = \infty$ otherwise. The latter one imposes a hard constraint on the total number of labels in the end result.

3.2. Co-occurrence Prior

To formulate the co-occurrence prior we make use of the functions l_i in (14). Let $L \subseteq \mathcal{G}$ denote a subset of labels which can appear in the current scene. For a given subset L let $\delta_L : \mathcal{S} \to \{0, 1\}$ denote the indicator function of a label subset $L \subseteq \mathcal{G}$ appearing in the scene :

$$\delta_L(u) = \begin{cases} 1 & \text{if } l_i(u) = 1 \; \forall i \in L, \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

Note that $\delta_L(u)$ is 1 if and only if every label from L occurs in image. Thus, δ_L is equal to:

$$\delta_L(u) = \prod_{i \in L} l_i(u), \tag{21}$$

and $E_L(l)$ can be written as:

$$E_L(l) = \sum_{L \subseteq \mathcal{G}} \delta_L(u) \cdot C(L) \tag{22}$$

with a label cost $C(L)$ associated with label configuration L. The term δ_L is not convex for $|L| \geq 2$. Thus a convex relaxation is required in order to convexify the final problem . Based on the tight convex relaxation given in [13] we obtain the following convex formulation :

$$E_L(l) = \sum_{L \subseteq \mathcal{G}} \left(\sup_{(\varphi_L, \psi_L) \in \mathcal{Q}_L} \sum_{i \in L} (1 - l_i)\varphi_L^i + l_i \psi_L^i \right) \tag{23}$$

We introduced dual variables $\varphi_L, \psi_L \in \mathbb{R}^{|L|}$ where $|L|$ stands for the cardinality of the subset L and where each dual variable pair φ_L^i, ψ_L^i is associated with a label indicator variable l_i for all $i \in L$. Additionally (φ_L, ψ_L) are constrained in the following set:

$$\begin{aligned} \mathcal{Q}_L := \Big\{ (\varphi_L, \psi_L) \Big| \; s.t. \; \forall z \in \{0,1\}^{|L|} \neq \vec{1} \quad (24) \\ \sum_{i \in L} (1 - z_i)\varphi_L^i + z_i \psi_L^i \leq 0 \text{ and } \sum_{i \in L} \psi_L^i \leq C(L) \Big\}, \end{aligned}$$

where $\vec{1}$ is a vector in $\mathbb{R}^{|L|}$ consisting of all ones. Since the power set of $\mathcal{G}$ is very large, as done in [8] we approximate the true costs by taking only sets with two labels into account with fixed costs, which best approximate the original cost function. The resulting formulation of $E_L(u)$ then reads as:

$$E_L(l) = \sum_{i<j \in \mathcal{G}} \left(\sup_{\{\varphi_{ij}, \psi_{ij}\} \in \mathcal{Q}_{i,j}} (1 - l_i)\varphi_{ij}^1 + l_i \psi_{ij}^1 \right. \tag{25}$$

$$\left. + (1 - l_j)\varphi_{ij}^2 + l_j \psi_{ij}^2 \right) \tag{26}$$

with $\mathcal{Q}_{ij}$ the convex constraint set of $\varphi_{ij}, \psi_{ij} \in \mathbb{R}^2$ given as follows:

$$\begin{aligned} \mathcal{Q}_{ij} := \Big\{ (\varphi_{ij}, \psi_{ij}) \Big| \text{ s.t. } \psi_{ij}^1 + \psi_{ij}^2 \leq C_{ij}, \quad (27) \\ \varphi_{ij}^1 + \varphi_{ij}^2 \leq 0, \; \varphi_{ij}^1 + \psi_{ij}^2 \leq 0, \; \psi_{ij}^1 + \varphi_{ij}^2 \leq 0 \Big\} \end{aligned}$$

3.3. Hierarchical Label Prior

Another application of the proposed general framework is the introduction of label hierarchies according to categories such as outdoor/indoor, animals, buildings etc., which consist of a set of labels $L \subset \mathcal{G}$. For each of these categories a specific cost $C(L)$ is defined. The hierarchical label prior is then defined as

$$E_L(l) = \sum_{L \subseteq \mathcal{G}} C(L) \max_{i \in L} l_i(u). \tag{28}$$

The above energy is non-differentiable and we derive a smooth convex optimization problem by use the following relaxation

$$\tilde{E}_L(l) = \sum_{L \subseteq \mathcal{G}} C(L) \left(\sup_{\mu_L} \sum_{i \in L} l_i(u)\, \mu_L^i \right) \tag{29}$$

$$\text{s.t.} \sum_{i \in L} \mu_L^i = 1,\ \mu_L^i \geq 0 \tag{30}$$

We introduced auxiliary dual variables $\mu_L \in \mathbb{R}^{|L|}$ where each μ^i is associated with the label indicator variable l_i for all $i \in L$. The equivalence of (28) and (29) is the well known dualization of the $\max$ function, see [15]. We give here a proof for the convenience of the reader:

Proof. We show the equivalence for two labels. The generalization to any larger set of labels is straight-forward. Let $l = (l_1, l_2)$ and binary, then

$$\tilde{E}_L(l) = \quad \sup_\mu (l_1(u)\mu_1 + l_2(u)\mu_2) \tag{31}$$

$$\text{s.t.} \quad \mu_1 + \mu_2 = 1 \tag{32}$$

$$\mu_1, \mu_2 \geq 0. \tag{33}$$

If $l_1(u) = l_2(u) = 0$ it is easy to see that the supremum over μ is 0 hence $\tilde{E}_L(l) = 0 = \max(l_1(u), l_2(u))$. In the case $(l_1(u) = 1$ or $l_2(u) = 1)$ the supremum is 1 and is reached for $\mu_1 = 1$ or respectively $\mu_2 = 1$. In the case $(l_1(u) = 1$ and $l_2(u) = 1)$ the supremum is 1 and is attained for $\mu_1 = \mu_2 = 0.5$ and the constraint (32) is active (32). Hence, we obtain $\tilde{E}_L(l) = E_L(l)$. □

4. Implementation

We use the primal-dual algorithm [2] which is essentially a gradient descent in the primal variables and a gradient ascent in the dual variables with a subsequent computation of the proximity operators. For the time steps we used the recent preconditioning techniques [12]. To impose the label configuration priors, three kinds of constraints have to be implemented.

- the **simplex constraint** $\sum_i u_i(x) = 1$ in (3) can be implemented by introducing Lagrange multipliers.

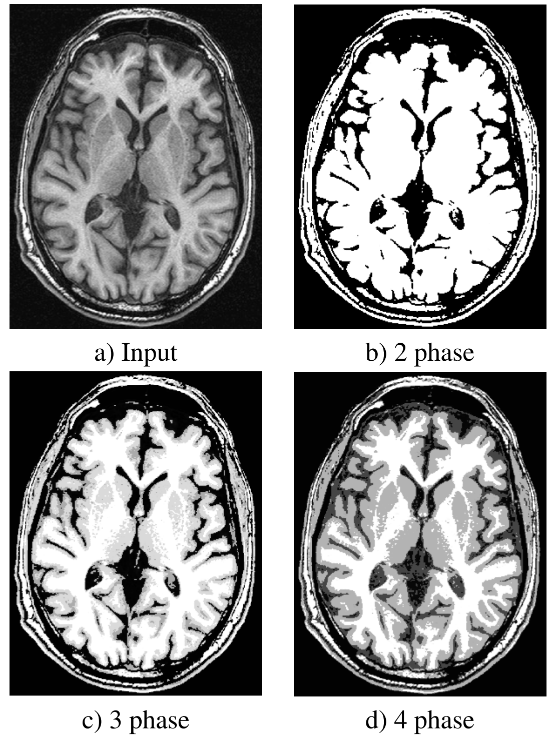


Figure 3. Generalized MDL prior: Minimizing energy (35) on an MRI image allows to impose an upper bound on the number of labels. This example shows an unsupervised segmentation of image (a) into 2 regions ($s_0 = 2$) (b), 3 regions ($s_0 = 3$) (c) and 4 regions ($s_0 = 4$) (d).

- **equality constraints**, e.g. $\sum_i \mu_{L_i} = 1$ in (29), are implemented by unconstrained Lagrange multipliers.
- **inequality constraints**, e.g. $l_i(u) \geq u_i(x)\ \forall x \in \Omega$, and the constraints in (27) are implemented by sign constrained Lagrange multipliers.

In addition we have the proximity operator of E_L in (19). In our case we use

$$f(s) = \delta_{s \leq s_0} \tag{34}$$

for some $s_0 \geq 1$ in Section 5.1. The resolvent operator for E_L is as follows:

$$l = \left(I + \tau \partial E_L\right)^{-1} (\tilde{l}) \Leftrightarrow l_i = \tilde{l} - \frac{1}{n} \max\left(s_0 - \sum_{i=1}^n \tilde{l}_i, 0\right)$$

5. Experiments

The key contribution of this paper is introducing label configuration priors into the continuous multi-label framework. We have applied the presented priors to various image segmentation tasks as will be shown in the following:

5.1. The Generalized MDL Prior

In the previous section we introduced the MDL prior and generalized it to convex, monotonously increasing functions f. We will show results for the case $f(s) = \delta_{s \leq s_0}$, i.e. for limiting the number of labels to at most s_0. We set the number of candidate labels to the number of gray levels $n = 256$ and the data term as the difference to a constant gray value for each region, $\varrho_i(x) = (I(x) - c_i)^2$ with $c_i = \frac{i}{255}$, $0 \leq i \leq 255$. We obtain the following optimization problem

$$\min_{\substack{u \in \mathcal{S} \\ l \text{ s.t. } (15)}} \max_{p \in K} \sum_{i=1}^{n} \int_{\Omega} u_i (I - c_i)^2 + u_i \operatorname{div} p_i \, dx. \tag{35}$$

with the set $\mathcal{S}$ in (12) and K in (7). This energy functional is a convex relaxation of the Mumford-Shah functional [11], which allows for the simultaneous optimization of the regions Ω_i and the gray values. We apply this model to the segmentation of an MRI-scan in Figure 3 and compute a 2, 3 and 4 phase segmentation of the input image. Note that the 4-phase segmentation gives best result since the white matter is separated properly from the gray matter.

5.2. Co-Occurrence

In this section we test the proposed continuous formulation of the segmentation problem with the co-occurrence prior on the MSRC database [5]. To preserve comparability to [8], we use their data term, which is based on texture boosting. Since the power set of the label set $\mathcal{G}$ is very large, the computation of the exact co-occurrence prior becomes intractable. Hence, as proposed in [8] we approximate based on subsets of only two labels. We applied the algorithm to the MSRC dataset. Some of the obtained segmentation results are shown in Figure 4. From the results we can conclude that the continuous formulation of the segmentation problem without superpixel preprocessing step improves on the results by Ladicky et al. [8] in several cases. The bird in the first row, for example, is segmented without its head by the approach in [8], since the head and the body are separated into two superpixels due to the strong color difference. The superpixels are not reunited by the segmentation approach. However, the continuous formulation, which is based on the single pixels, is able to recover the head.

A comparison of the segmentation accuracy of the proposed algorithm to the original formulation by Ladicky et al. on the whole benchmark can be found in Figure 5. The table indicates the numbers for each label separately as well as the average on the whole benchmark. The results show that the segmentations obtained with the proposed continuous formulation are comparable to those obtained by Ladicky et al. in terms of overall accuracy, and even outperforms them in terms of average accuracy. Note that in contrast to Ladicky et al. we do not use any preprocessing steps such as superpixel extraction.

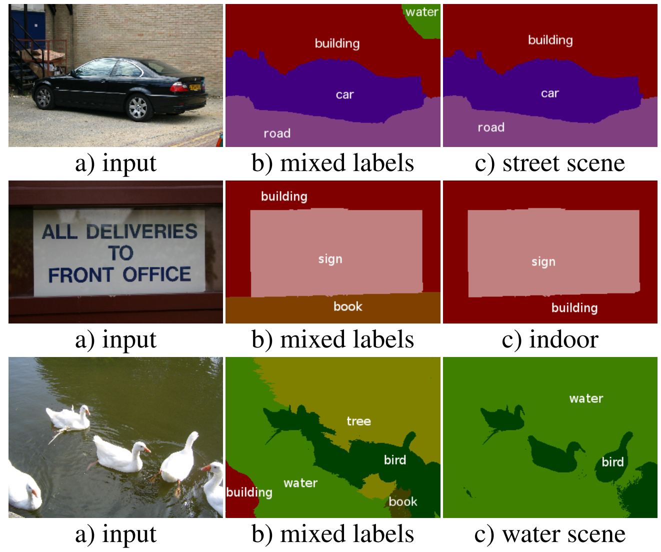


Figure 7. Hierarchical Prior: A joint semantic segmentation and scene classification can be computed using hierarchical priors as shown in Figure 6. Such scene-adaptive priors allow to avoid scene-inconsistent semantic segmentations.

5.3. Hierarchical Label Prior

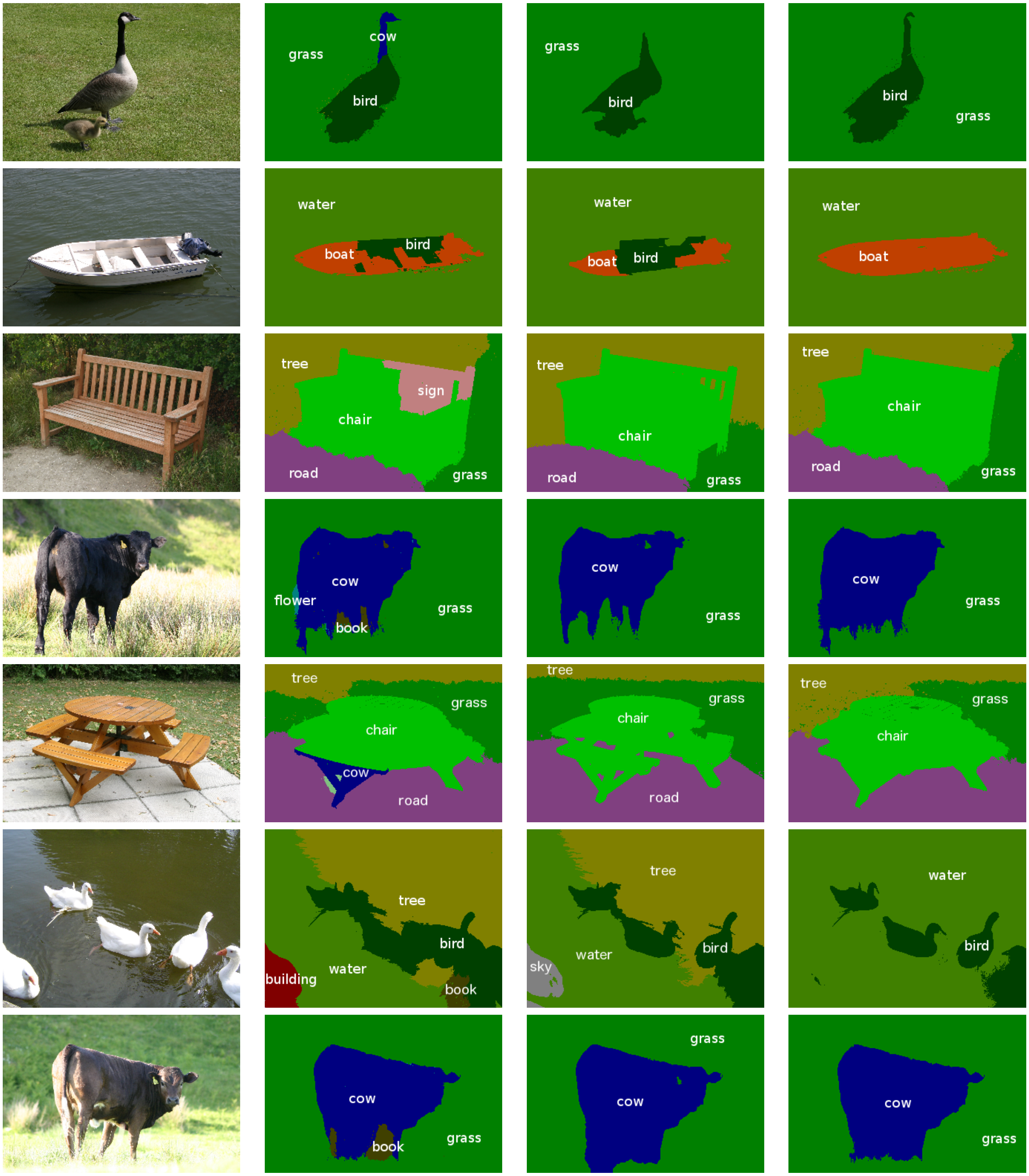
In the following we will demonstrate that using the hierarchical prior introduced in Section 3.3 we can joinly incorporate a semantic prior and infere nature of the scene at hand. For this we apply our algorithm on images from the MSRC database and the following label set

$$\mathcal{G} := \{\text{Car, Bird, Building, Sky, Water, Boat,} \\ \text{Chair, Tree, Sign, Road, Book}\}.$$

Next we partition the label space $\mathcal{G}$ into 3 Subgroups *Street Scene*, *See Scene* and *Indoor Scene*. The group membership of the labels in each group is illustrated in Figure 6. We define the costs $C(\text{Street Scene})$, $C(\text{Water Scene})$, $C(\text{Indoor Scene})$ for choosing one or more labels from the respective category. The results of our optimization algorithm can be seen in Figure 7.

6. Runtime

The proposed algorithm is based on a variational approach which allows for an implementation on graphics hardware. Furthermore, we do not use any preprocessing step such as the computation of superpixels. For the experiments we used a NVIDIA Geforce GTX480 GPU and obtained average runtimes per image of 10 seconds for the co-occurrence segmentation as well as for the hierarchical prior segmentation. The runtimes are similar, since the



a) Input b) Potts c) Ladicky et al. [8] d) Configuration Prior

Figure 4. Co-occurrence prior: Qualitative results on images taken from the MSRC database. The results show (a) the original benchmark image, (b) the segmentation result without co-occurrence prior, (c) the results by Ladicky et al. [8], (d) results from our continuous formulation of the segmentation algorithm incorporating the co-occurrence energy.

	Global	Average	Building	Grass	Tree	Cow	Sheep	Sky	Aeroplane	Water	Face	Car	Bicycle	Flower	Sign	Bird	Book	Chair	Road	Cat	Dog	Body	Boat
dataterm	83.99	77.18	67	97	91	85	86	95	88	81	90	82	94	81	62	42	91	66	86	79	54	72	31
potts	84.80	77.88	69	97	91	86	86	96	86	82	90	81	93	83	62	42	91	68	86	80	56	72	28
proposed	85.99	**78.95**	72	97	91	87	86	97	86	84	90	83	93	83	64	44	93	73	87	81	59	74	25
Ladicky et al. [8]	**86.76**	77.78	76	99	90	77	84	99	82	88	88	80	90	90	71	47	94	68	90	73	55	77	15

Figure 5. Segmentation accuracies of the daterterm, the pure Potts model, our approach using and the resuls of Ladicky et al. [8]. The scores for each label are defined as $\frac{\textit{True Positives} \cdot 100}{\textit{True Positives} + \textit{False Negatives}}$. While the proposed method is worse than [8] in the global score, it provides a better average performance.

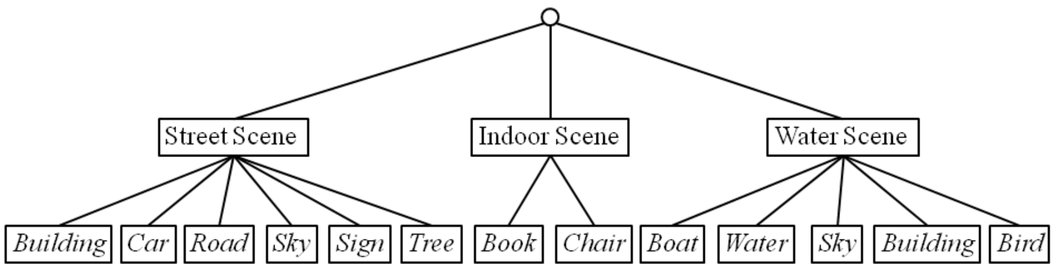


Figure 6. As an instance of a configuration prior we can impose a **hierarchal prior** which leads to a scene parsing which jointly estimates a scene classification and a semantic partitioning of the image – see Figure 7.

overhead of computation compared to the classical multi-labeling problems is marginal as the configuration priors are defined on n scalar indicator variables compared to $\mathcal{O}(|\Omega|)$ variable for the indicator functions u.

7. Conclusion

We propose a convex framework for multi-label optimization which allows to incorporate label configuration priors such as generalized MDL priors, co-occurrence priors and hierarchical label cost priors. To this end, we formulate a variational approach which can be optimized with fast primal-dual schemes. In contrast to existing spatially discrete MRF-based approaches, the proposed method does not require any pre-processing in terms of super-pixel estimation but directly works on the input pixels.

References

[1] A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Technical report TR-2008-05, Dept. of Computer Science, University of Bonn, Bonn, Germany, Nov. 2008. 2

[2] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *JMIV*, 40(1):120–145, 2011. 5

[3] A. Delong, L. Gorelick, and Y. Boykov. Minimizing energies with hierarchical costs. *IJCV*, 2012. To appear. 1, 2

[4] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. *IJCV*, 96(1):1–27, 2012. 1, 2, 4

[5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. Pascal (VOC2010) Results. 6

[6] B. Goldluecke and D. Cremers. Convex relaxation for multilabel problems with product label spaces. In *ECCV*, 2010. 2

[7] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *ECCV*, Marseille, France, October 2008. 2

[8] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr. Graph cut based inference with co-occurrence statistics. In *ECCV*, pages 239–253, 2010. 1, 2, 3, 4, 6, 7, 8

[9] Y. G. Leclerc. Region growing using the MDL principle. In *Proc. DARPA Image Underst. Workshop*, pages 720–726, April 6-8 1990. 1

[10] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. Technical report, October 2008. 2

[11] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989. 6

[12] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *ICCV*, 2011. To appear. 5

[13] E. Strekalovskiy, B. Goldluecke, and D. Cremers. Tight convex relaxations for vector-valued labeling problems. In *IEEE International Conference on Computer Vision (ICCV)*, 2011. 2, 4

[14] M. Werlberger, M. Unger, T. Pock, and H. Bischof. Efficient Minimization of the Non-Local Potts Model. In *ICSSVMCV*, 2011. 2, 3

[15] J. Yuan and Y. Boykov. Tv-based multi-label image segmentation with label cost prior. In *BMVC*, pages 1–12, 2010. 1, 2, 4, 5

[16] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. October 2008. 2

[17] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multi-band image segmentation. *PAMI*, 18:884–900, 1996. 1