Nonmetric Priors for Continuous Multilabel Optimization

Evgeny Strekalovskiy, Claudia Nieuwenhuis, and Daniel Cremers

Technical University Munich, Germany

Abstract. We propose a novel convex prior for multilabel optimization which allows to impose arbitrary distances between labels. Only symmetry, $d(i, j) \geq 0$ and $d(i, i) = 0$ are required. In contrast to previous grid based approaches for the nonmetric case, the proposed prior is formulated in the continuous setting avoiding grid artifacts. In particular, the model is easy to implement, provides a convex relaxation for the Mumford-Shah functional and yields comparable or superior results on the MSRC segmentation database comparing to metric or grid based approaches.

1 Introduction

1.1 Nonmetric Priors in Image Segmentation

The task of multilabel optimization is to assign a meaningful *label* to each image pixel so that the overall labeling minimizes a certain cost. Example problems include 3D-reconstruction, optical flow and denoising. Another one is segmentation where the image is to be divided into a set of non-overlapping regions, which are homogeneous in specific way. Respective algorithms usually define color models for each object or region, e.g. by estimating probability distributions in the color space. Adding *prior* constraints such as minimal boundary or curvature of the segmented objects, impressive results can be obtained. Yet, for a large number of objects additional information is indispensable to resolve the ambiguity between objects of similar or mixed colors. For example, cows, sheep and horses all contain colors of white, black and brown. Depending on the color distributions, this often leads to mixed animals, having e.g. 'cow' labeled bodies and 'sheep' labeled heads. To avoid such problems, label transitions can be punished depending on the probability of co-occurrence of two different objects next to each other. This, however, in general leads to *nonmetric* arbitrary label distance functions, which cannot be handled by common algorithms such as α-expansion [1] or primal dual schemes [2,3], which assume metrical distance functions.

Nonmetric priors are encountered in many practical problems, since triangle inequality is usually not preserved and infinite distances appear between entirely unrelated objects. For example, while sheep and tigers are both frequently encountered in grass, nevertheless they usually do not appear next to each other. A challenge which has largely been neglected is to devise algorithms which allow to impose such nonmetric distance priors into multilabel optimization to be able to apply any learned co-occurrence relations.

1.2 Related Work

Multilabel problems are generally solved by minimizing energies of the form

$$E(u) = E_{\text{data}}(u) + E_{\text{reg}}(u). \tag{1.1}$$

The data term E_{data} corresponds to the negative log likelihood for observing the data given a specific model such as a learned color model, whereas the regularizer E_{reg} corresponds to the negative logarithm of some prior stating which label configurations are apriori more likely. For reasons of invariance to translation in the label space, priors are typically functions of the distance between label values assigned to neighboring nodes.

A simple prior is the linear label cost, i.e. for two labels i and j and a constant c the label distance function can be expressed by $d(i, j) = c\, |i - j|$. In the MRF domain, the corresponding labeling problems can be optimized globally by the Ishikawa approach [4]. In the continuous setting, this formulation corresponds to the total variation regularization.

In the case of a metric distance function, in general no globally optimal solutions can be obtained for more than two labels. A popular metric prior is the Pott's model [5], which assigns a constant cost to each label change independent of the labels. In the MRF domain, algorithms such as α-expansion and α-β swap [1] were developed to solve the corresponding multilabel problem. In the continuous setting, to this end convex relaxation schemes were proposed by Chambolle et al. [2], Pock et al. [6] and Lellmann et al. [3,7,8].

In [7], it was shown how metrical approximations to nonmetric distance functions can be obtained. In practice, however, these approximations can be arbitrarily far from the original distance function, e.g. in case of learned distance functions. Guaranteed integrality gaps between the relaxed and the integer solution for different approximations to the metric labeling problem have been proven by Chekuri et al. [9].

In the specific field of multilabel segmentation, nonmetric distance functions have been introduced before. Geodesic distances have been formulated by Bai et al. [10]. Co-occurrence probabilities, which penalize the simultaneous appearance of label sets within an image and thus implicitly influence neighboring labels, have been modeled by Ladicky et al. [11].

Discrete Nonmetric Labeling. Most closely related to our work is the approach by Chekuri et al. [9] for handling general nonmetric distance functions in the MRF domain. It is able to handle arbitrary *label distance functions*

$$d : \{1, \ldots, n\}^2 \to \mathbb{R} \tag{1.2}$$

where $n \geq 1$ is the number of labels. In particular, neither symmetry nor the triangle inequality are assumed on the label distances. The distance $d(i, j)$ gives the penalization if the multilabel assignment changes from label i to label j.

In the discrete case of MRF energies defined on a graph (V, E), it gives an LP-relaxation for the general potentials of second order. The regularizer is defined

separately on each edge of the graph and gives a penalization if the labeling $u : V \to \mathcal{L} := \{1, \ldots, n\}$ is different on the two endpoints of the edge:

$$R(u) = \sum_{(a,b) \in E} d(u(a), u(b)) \,. \qquad (1.3)$$

For each label i a label indicator function $u_i(a) \in \{0,1\}$ is introduced, with $u_i(a) = 1$ if the label i is set in pixel a and $u_i(a) = 0$ otherwise. The regularizer part of the LP-relaxation in [9] is

$$R(u) = \inf_{u_{ij} \geq 0} \sum_{(a,b) \in E} \sum_{i,j \in \mathcal{L}} d(i,j) u_{ij}(a,b) \qquad (1.4)$$

where the new variables $u_{ij}(a,b)$ are constrained by

$$
\begin{aligned}
\sum_j u_{ij}(a,b) &= u_i(a) \quad \forall i \in \mathcal{L}, \\
\sum_i u_{ij}(a,b) &= u_j(b) \quad \forall j \in \mathcal{L}.
\end{aligned}
\qquad (1.5)
$$

However, because of the MRF domain the regularizer (1.4) is defined directly on the underlying pixel grid. Therefore, the interface length between two labels is measured in the l^1 instead of the l^2 norm leading to grid bias, i.e. grid aligned interfaces between labels are favored.

In contrast, in the continuous setting, variational multilabel approaches have so far been limited to *metric* distance measures leading to crude approximations of the actual label distances.

1.3   Contribution

In this paper, we present a novel spatially continuous approach to the multilabel problem, which allows for arbitrary label distances. We formulate an efficient primal-dual algorithm and compare results to previous approaches, which are restricted to metric label distances, on the MSRC segmentation benchmark.

More specifically, our contributions are as follows:

- We propose a novel regularizer for multilabel optimization which can handle arbitrary explicitly specified label distances.
- The regularizer is spatially continuous and therefore is rotationally invariant, avoiding the grid bias.
- The model is easy to implement and the results are comparable and often superior to the metric and grid based approaches.

2   Continuous Multilabel Optimization with Nonmetric Priors

The general multilabel problem in image domain $\Omega \subset \mathbb{R}^2$ with $n \geq 1$ labels consists in finding n label indicator functions $u_1, \ldots, u_n : \Omega \to \{0, 1\}$ minimizing

$$\inf_u \left\{ \sum_{i=1}^n \int_\Omega \varrho_i(x) u_i(x)\, dx \; + \; R(u) \right\} \tag{2.1}$$

under the label-uniqueness constraint $\sum_{i=1}^n u_i(x) = 1$. Here, $\varrho_i(x)$ is the data term, i.e. the local cost of assigning label i at image point $x \in \Omega$, and $R(u)$ a multilabel *regularizer* ensuring a spatial consistency of labels.

Ideally, we want the regularizer $R(u)$ to measure the total length of all label interfaces, weighted by the distances $d(i, j)$ of the corresponding labels i and j at the two interface sides. In the following, we propose a definition of such a regularizer and show in Theorem 1 that it indeed satisfies this requirement.

2.1   The Novel Regularizer

We propose the following regularizer:

$$R(u) = \sup_{(p,q) \in C} \sum_{i=1}^n \int_\Omega \Big(p_i(x) \nabla u_i(x) + q_i(x) u_i(x) \Big)\, dx \tag{2.2}$$

with the convex set

$$C = \Big\{ (p, q) : \Omega \to (\mathbb{R}^2 \times \mathbb{R})^n \;\big|\; |p_j(x) - p_i(x)| + q_i(x) \leq d(i, j) \quad \forall x, i, j \Big\}. \tag{2.3}$$

The dual variables consist of n vector fields $p_1, \ldots, p_n$ and n scalars $q_1, \ldots, q_n$, corresponding to the n labels $u_1, \ldots, u_n$. The distance d may be *arbitrary*. In particular we do not require it to be a metric. As usual, the distance measure notion implies symmetry, $d(i, j) \geq 0$ for all i, j, and $d(i, i) = 0$ for all i. These are the only conditions we impose on d.

Since $u : \Omega \to \{0, 1\}^n$ is in general non-smooth, we write the term $p_i \nabla u_i$ only for convenience, tacitly meaning $(-\operatorname{div} p_i) u_i$ and assuming p to be smooth.

2.2   Motivation for the Definition

Introducing Lagrange multipliers $v_i(a, b), w_j(a, b) \in \mathbb{R}$ for the constraints (1.5), the expression (1.4) can be written as

$$R_{(a,b)} = \sup_{v,w} \sum_{(a,b) \in E} \left(\sum_i v_i(a, b) u_i(a) + \sum_j w_j(a, b) u_j(b) \right) \tag{2.4}$$

with v, w such that

$$v_i(a, b) + w_j(a, b) \leq d(i, j) \quad \forall i, j. \tag{2.5}$$

Writing $\widehat{p}_i(a,b) := w_i(a,b)$ and $\widehat{q}_i(a,b) := v_i(a,b) + w_i(a,b)$ this becomes

$$R_{(a,b)} = \sup_{\widehat{p},\widehat{q}} \sum_{(a,b)\in E} \sum_i \left(\widehat{p}_i(a,b)\big(u_i(b) - u_i(a)\big) + \widehat{q}_i(a,b)\,u_i(a) \right) \qquad (2.6)$$

with the constraints

$$\big(\widehat{p}_j(a,b) - \widehat{p}_i(a,b)\big) + \widehat{q}_i(a,b) \leq d(i,j) \quad \forall i,j. \qquad (2.7)$$

This expression is already very similar to the proposed one (2.2). We can replace the sum over the edges in (2.6) by the sum over each vertex $a \in V$, considering its right and upper neighbors b_h and b_v, respectively:

$$R_{(a,b)} = \sup_{\widehat{p},\widehat{q}} \sum_{a\in V} \sum_i \left(\begin{pmatrix} \widehat{p}_i(a,b_h) \\ \widehat{p}_i(a,b_v) \end{pmatrix} \begin{pmatrix} u_i(b_h) - u_i(a) \\ u_i(b_v) - u_i(a) \end{pmatrix} + \big(\widehat{q}_i(a,b_h) + \widehat{q}_i(a,b_v)\big)\,u_i(a) \right).$$

Note that the vectorial expression in u_i can be regarded as the discretization of the gradient ∇u_i at $a \in V$. The crucial step to arrive at (2.2) is now to define

$$p_i(a) := \begin{pmatrix} \widehat{p}_i(a,b_h) \\ \widehat{p}_i(a,b_v) \end{pmatrix} \quad \text{and} \quad q_i(a) := \widehat{q}_i(a,b_h) + \widehat{q}_i(a,b_v) \qquad (2.8)$$

and to replace the two constraints (2.7) for $b = b_h$ and $b = b_v$ by one:

$$|p_j(a) - p_i(a)| + q_i(a) \leq d(i,j) \quad \forall i,j. \qquad (2.9)$$

2.3   Properties of the Regularizer

We prove the following main theorem of the paper. It shows that while the constraints (2.9) are not equivalent to (2.7) they give rise to a rotationally invariant regularizer having the desired penalization properties.

Theorem 1. *Let $u = e_i\chi_A + e_j\chi_{\bar{A}}$ with a subset $A \subset \Omega$ and some fixed $i,j \in \{1,\ldots,n\}$. Then*

$$R(u) = d(i,j)\,\mathrm{Per}(A) \qquad (2.10)$$

where $\mathrm{Per}(A) = TV(\chi_A)$ is the perimeter of A in Ω.

Proof. See appendix.

In other words, a labeling change from label i to label j will be penalized by the label distance $d(i,j)$ weighted by the length of the interface between the regions where these labels are attained. Furthermore, the regularizer $R(u)$ has the favorable property of being convex, rendering global optimization possible.

Proposition 1. *$R(u)$ is convex and lower-semicontinuous on $L^2(\Omega;\mathbb{R}^n)$.*

Proof. From (2.2) it follows that $R(u)$ is the support functional

$$R(u) = \sup_{z\in D} \sum_{i=1}^n \int_\Omega u_i(x) \cdot z_i(x)\,dx \qquad (2.11)$$

of the convex set $D = \{(-\operatorname{div} p_i + q_i)_{i=1,\ldots,n} \,\big|\, (p,q) \in C\}$ with the set C in (2.3). Thus, the theorem follows directly from the properties of support functionals [12, Theorem 9.1.2].

2.4   Special Case: Regularizer for Metrics

For the special case $q_i(x) \equiv 0$, the proposed regularizer reduces to the known regularizer of [2] for *metric* distances d, namely

$$R_m(u) = \sup_{p \in C_m} \sum_{i=1}^{n} \int_{\Omega} p_i(x) \nabla u_i(x) \, dx \qquad (2.12)$$

with the convex set

$$C_m = \Big\{ p : \Omega \to (\mathbb{R}^2)^n \ \big| \ |p_j(x) - p_i(x)| \le d(i,j) \quad \forall i,j \Big\}. \qquad (2.13)$$

Note that R_m is still applicable also for nonmetric distances d. In that case, due to the definition of the set C_m it will implicitly work with a "truncated" version of d, namely the metric

$$\widehat{d}(i,j) := \sup_{p \in \mathbb{R}^n : \, |p_l - p_k| \le d(k,l) \, \forall k,l} |p_j - p_i|. \qquad (2.14)$$

Lellmann et al. [3] proposed a regularizer in the continuous domain for metric distances d only, and showed in Proposition 3.1 that d must necessarily be a metric if the regularizer R satisfies certain simple conditions. Since our proposed regularizer (2.2) can handle arbitrary nonmetric distances by Theorem 1, it is interesting to see which of these conditions are *not* satisfied in our case. It turns out to be only the second condition (P2) which states that the regularizer must be zero for any constant $u : \Omega \to \mathbb{R}^n$. In fact, we have the following result:

Proposition 2. *Let $u : \Omega \to \mathbb{R}^n$ be constant, $u_i(x) = c_i$ for all i and $x \in \Omega$ with some $c \in \mathbb{R}^n$. Then $R(u) = 0$ if $c_i \ge 0$ for all i, and $R(u) = \infty$ otherwise.*

Proof. See appendix.

Because in the optimization we have $u_i(x) \in \{0,1\}$ (respectively $u_i(x) \in [0,1]$ after relaxation) the proposed regularizer *is* zero for any constant u which represents a labeling. Thus, dropping the condition (P2) of [3] for non-meaningful labelings allows us to handle arbitrary distances, and not only the metric ones.

2.5   Convex optimization

The problem (2.1) is not convex because of the nonconvex constraints $u_i(x) \in \{0,1\}$. To obtain a convex optimization problem, we relax this binary constraint to $u_i(x) \in [0,1]$ for all labels i. Thus, the minimization is performed over the candidate set

$$K := \Big\{ u \in L^2(\Omega; \mathbb{R}^n) \ \big| \ u_i(x) \in [0,1], \ \sum_{j=1}^{n} u_j(x) = 1 \ \ \forall i, x \in \Omega \Big\} \qquad (2.15)$$

One can easily prove the existence of minimizers for the overall optimization problem (2.1) for general dataterms $\varrho \in L^2(\Omega; \mathbb{R}^n)$.

Theorem 2. *The problem (2.1), (2.2) admits a minimizer $u \in K$.*

Proof. See appendix.

3 Implementation

We solve the overall optimization problem (2.1), (2.2) using the general algorithm of [13]. This is a fast primal-dual algorithm which consists essentially in a gradient descent in u and a gradient ascent in p, with an orthogonal reprojection onto the constraint set C in (2.3). The individual constraints of C are implemented using convex duality. To this end, we add the terms

$$\inf_{|a_{ij}| \le b_{ij}} \sum_{i,j} \int_\Omega \big(- a_{ij}(p_j - p_i) - b_{ij}q_i + b_{ij}d(i,j) \big)\, dx \qquad (3.1)$$

to the energy and optimize the overall problem also in a and b, with no constraints on p, q. The simplex constraint $\sum_i u_i(x) = 1$ for each $x \in \Omega$ is also implemented using Lagrange multipliers. For this, we replace it by $u_i(x) \ge 0$ and add the Lagrange multiplier terms

$$\sup_\sigma \int_\Omega \sigma(x) \left(\sum_{i=1}^n u_i(x) - 1 \right) dx \qquad (3.2)$$

to the energy and optimize over $\sigma : \Omega \to \mathbb{R}$. In comparison to the direct projection, this leads to significantly faster updates per iteration, requiring only slightly more overall iterations. The time steps for the algorithm are chosen automatically using the recent preconditioning scheme [14]. We used the same algorithm [13] to implement the discrete model [9].

4 Experiments

We used 5000 iterations for each experiment after which the solutions become visually stable. We used a parallel CUDA implementation on NVIDIA GTX 480. Usual runtimes for 320×240 images and 21 labels are around 70 seconds. We observed that the computed relaxed solutions u are binary almost everywhere except at region boundaries, with more of less sharp transitions. We binarize the result at each pixel x by taking the label i with the maximal value $u_i(x)$.

4.1 Piecewise Smooth Mumford-Shah Functional

First we demonstrate the application of our approach on the celebrated Mumford-Shah functional [15,16]. In the continuous domain it is given by

$$E(u, K) = \int_\Omega (u - f)^2\, dx + \alpha \int_{\Omega \backslash K} |\nabla u|^2\, dx + \nu\, |K| \qquad (4.1)$$

Given a possibly noisy input image $f : \Omega \to \mathbb{R}$ this yields *piecewise smooth* approximations $u : \Omega \to \mathbb{R}$ of f. The function u will be smooth, except possibly for a one-dimensional edge set K where jumps occur. The parameter ν controls

| Input | discrete [9] | continuous [16] | proposed, same as with [16] |
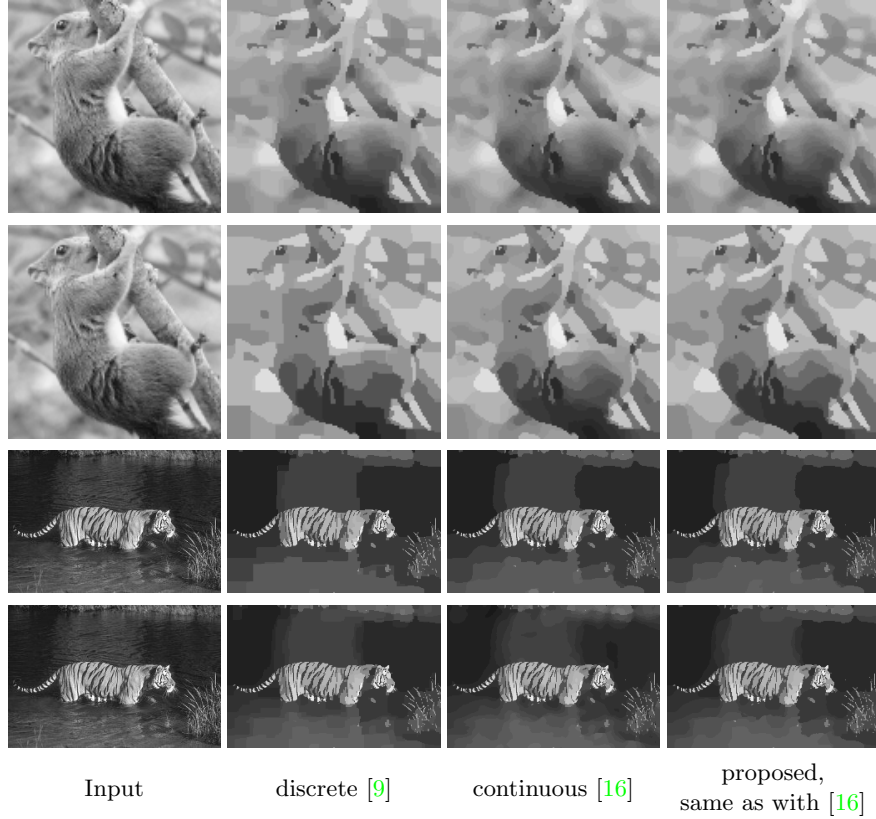
Fig. 1. Piecewise smooth approximations. Although [9] can handle the nonmetric label distance function induces by the Mumford-Shah functional, it favors grid aligned edges producing block artifacts. In contrast, the proposed approach is based in the continuous setting and produces results visually indistinguishable from [16]. While the nonmetric is implicit in [16], we can specify any nonmetric in a direct way.

the length of the jump set K. Bigger values of ν lead to a smaller jump set, i.e. the solution will be smooth on wider subregions of Ω.

Discretizing the range $[0, 1]$ of u into n labels, this leads to a multilabel problem with the well-known *truncated quadratic* label distances:
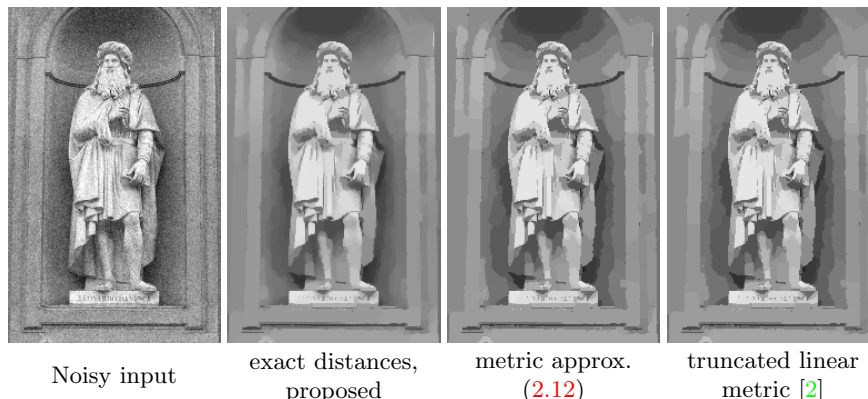
$$d_{MS}(i, j) = \min\left(\nu, \, \alpha\left(\tfrac{i-j}{n}\right)^2\right) \qquad \forall 1 \le i, j \le n. \tag{4.2}$$

This distance function is a not metric for $\nu > \frac{2\alpha}{n^2}$. For example,

$$d_{MS}(0, 2) = \min(\nu, \tfrac{4\alpha}{n^2}) > 2\min(\nu, \tfrac{\alpha}{n^2}) = d_{MS}(0, 1) + d_{MS}(1, 2). \tag{4.3}$$

Therefore, our approach applies naturally here. Some results for different parameters ν and α are shown in Figure 1. As expected, our approach produces

| Noisy input | exact distances, proposed | metric approx. (2.12) | truncated linear metric [2] |

Fig. 2. Nonmetric versus metric approaches. From left to right: A noisy input image. Piecewise smooth approximation using our approach by specifying the nonmetric Mumford-Shah label distances. Ignoring the nonmetric character of the distance function and treating it as a metric by (2.12) effectively imposes a "truncated linear" metric, which leads to staircasing effects. This is further confirmed by solving for this metric explicitly using [2].

visually the same results as Pock et al. [16] (handles exclusively the Mumford-Shah model) since both work in the continuous setting. While their approach uses advanced tools such as functional lifting to arrive at the convex relaxation, the proposed approach is more basic as it specifies the label distances explicitly. The discrete grid based approach [9] can also handle the nonmetric distance (4.2) but measures the interface length in the l^1 norm. As seen in Fig. 1, our model evidently visually improves over [9], eliminating its block artifacts.

The experiment in Figure 2 shows the importance of the ability of the proposed approach to handle nonmetric distances exactly. Trying to solve the Mumford-Shah problem using the metric approximation (2.12) effectively imposes the *truncated linear* prior instead of truncated quadratic, which leads to staircasing effects. In fact, one easily show that the truncated metric in (2.14) is

$$\widehat{d}_{MS}(i,j) = \min\left(\nu,\ \alpha\,\frac{|i-j|}{n^2}\right). \tag{4.4}$$

4.2   MSRC Segmentation Benchmark

To evaluate the proposed segmentation algorithm we apply it to the task of object segmentation and recognition on the MSRC benchmark. This benchmark comprises around 600 images which contain 23 different labels such as cow, book, building or grass. To conduct experiments on this benchmark, we follow Ladicky et al. [11] and divide the image set randomly into 60% training images and 40% test images.

The label distance matrix is learned on the training set. For each pair of labels (i,j), $i < j$ we compute the relative frequency that i and j are neighbors
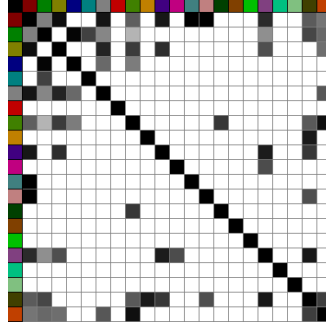
Fig. 3. Label distance matrix for the MSRC database in Section 4.2. The distance function is estimated on the training set and is an example of a *nonmetric*. The first row and column are the color coded labels and the other entries depict the values $d(i,j)$ with 0 as black and 10 as white.

in the following way. Let A_{ij} and B_{ij} denote the number of pairs of pixels where labels i and j are direct or diagonal neighbors, respectively:

$$A_{ij} := \#\{(x,y) \in \Omega \times \Omega \mid l(x) = i,\, l(y) = j,\, \|x-y\|_2 = 1\},$$
$$B_{ij} := \#\{(x,y) \in \Omega \times \Omega \mid l(x) = i,\, l(y) = j,\, \|x-y\|_2 = \sqrt{2}\,\}. \tag{4.5}$$

Then the label distance is calculated as the negative logarithm of the relative weighted frequency:

$$d(i,j) = -\log \frac{A_{ij} + \frac{B_{ij}}{\sqrt{2}}}{\sum_k \left(A_{ik} + \frac{B_{ik}}{\sqrt{2}}\right)} \, . \tag{4.6}$$

For non-adjacent labels i,j with $|A| = |B| = 0$, we truncate $d(i,j) = \infty$ to $d(i,j) = M$ for some $M > 0$ (e.g. 10). Figure 3 indicates the distance function obtained from the MSRC training set. The brightness of the entry corresponds to the distance of the coinciding labels.

To evaluate the segmentation accuracy of the proposed method, we compare the scores and the overall accuracies for the following approaches:

- the proposed nonmetric regularization (2.2),
- the "truncation to metric" regularization (2.12),
- the discrete l^1 nonmetric regularization (1.4),
- the co-occurrence statistics based approach by Ladicky et al. [11].

Nonmetric regularization means the approach proposed in this paper, which can handle arbitrary distance functions. Truncation to metric means that the learned distance function d is approximated by the closest metrical distance function in the continuous domain. The l^1 nonmetric regularization penalizes distances in horizontal and vertical direction separately leading to a direction

| Input | Metric (2.12) | Discrete [9] | Ladicky et al. | Proposed |
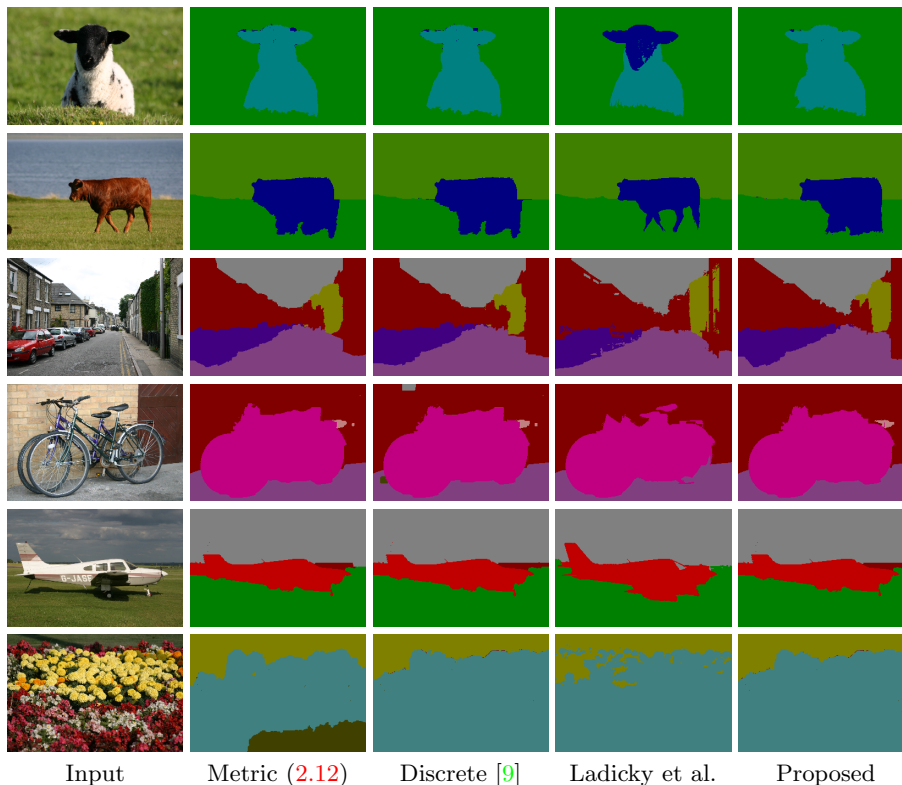
Fig. 4. Results on the MSRC segmentation benchmark. *From left to right: Input image, segmentation by* (2.12) *truncating the label distance to the nearest metric, segmentation using the grid based approach* (1.4), *segmentation using the proposed regularizer* (2.2). *The regularizer weighting is optimized separately for each approach.*

dependent distance function. In the discrete case, this formulation corresponds to Chekuri et al. [9]. Finally, we compare the obtained benchmark results to those reported by Ladicky et al. [11]. In contrast to the proposed method which defines distance based on neighboring pixel labels (second order potentials), the authors of [11] use information on the general co-occurrence of two labels in one image to derive label distances (potentials of the highest order $|\Omega|$).

Results for the different approaches are shown in Figure 4. A quantitative comparison for each label as well as the overall accuracies can be seen in Table 1. We set the regularizer weighting to $\nu = 0.2$ for every image in the database. The proposed approach leads to best overall accuracies comparing with the metric and grid based approaches.

	Accuracy	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
metric	84.72	77.54	70	97	91	88	86	96	83	82	90	82	93	83	65	46	92	64	87	81	51	73	31
discrete l^1	84.79	**77.62**	69	97	91	88	86	96	83	83	90	83	93	83	64	46	92	63	86	80	51	73	31
proposed	**84.82**	**77.62**	69	97	92	88	86	96	83	83	90	82	93	83	64	46	92	64	86	80	51	73	32
Ladicky et al. [11]	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

Table 1. Segmentation accuracies for the different approaches. The scores for each label are defined as $\frac{True\ Positives}{True\ Positives\ +\ False\ Negatives}$. Also shown is the average over the scores and the accuracy for each approach, i.e. the overall number of pixels labeled correctly as in the ground truth, divided by the overall number of pixels. The proposed continuous nonmetric approach provides the best overall accuracy. Results of [11] are only for comparison, since they use potentials of the highest order $|\Omega|$, instead of order two as in our approach.

5   Conclusion

For the multilabel optimization problem we introduced a novel regularizer which can handle arbitrary label distances. In contrast to previous discrete approaches, it is based in the continuous setting and does not suffer from metrical artifacts. Being convex it allows to find globally optimal solutions of the relaxation. The proposed model leads to consistently better results than the discrete model. Experimental results show competitiveness to state-of-the-art discrete approaches. On the MSRC segmentation database we get higher overall accuracy, and the Mumford-Shah experiments evidently show visual improvements over the discrete model, eliminating its block artifacts. Future applications, with a more distinct nonmetric structure of the distance functions based on our approach may lead to substantial improvements.

References

1. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. on Patt. Anal. and Mach. Intell. **23**(11) (2001) 1222–1239 1, 2
2. Chambolle, A., Cremers, D., Pock, T.: A convex approach for computing minimal partitions. Technical report TR-2008-05, Dept. of Computer Science, University of Bonn, Bonn, Germany (November 2008) 1, 2, 6, 9
3. Lellmann, J., Schnörr, C.: Continuous Multiclass Labeling Approaches and Algorithms. SIAM J. Imag. Sci. **4**(4) (2011) 1049–1096 1, 2, 6
4. Ishikawa, H.: Exact optimization for Markov random fields with convex priors. IEEE Trans. on Patt. Anal. and Mach. Intell. **25**(10) (October 2003) 1333–1336 2
5. Potts, R.B.: Some generalized order-disorder transformations. Proc. Camb. Phil. Soc. **48** (1952) 106–109 2
6. Pock, T., Chambolle, A., Bischof, H., Cremers, D.: A convex relaxation approach for computing minimal partitions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, Florida (2009) 2

7. Lellmann, J., Becker, F., Schnörr, C.: Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In: IEEE Int. Conf. on Computer Vision. (2009) 646 – 653 2
8. Lellmann, J., Breitenreicher, D., Schnörr, C.: Fast and exact primal-dual iterations for variational problems in computer vision. In: Europ. Conf. on Computer Vision. (2010) 2
9. Chekuri, C., Khanna, S., Naor, J., Zosin, L.: A linear programming formulation and approximation algorithms for the metric labeling problem. SIAM Journal on Discrete Mathematics **18** (2005) 608–625 2, 3, 7, 8, 9, 11
10. Bai, X., Sapiro, G.: A geodesic framework for fast interactive image and video segmentation and matting. In: IEEE Int. Conf. on Computer Vision. (2007) 2
11. Ladicky, L., Russell, C., Kohli, P., Torr, P.: Inference methods for crfs with co-occurence statistics. In: Int. J. of Computer Vision. (2010) 2, 9, 10, 11, 12
12. Attouch, H., Buttazzo, G., Michaille, G.: Variational Analysis in Sobolev and BV Spaces. MPS-SIAM Series on Optimization. SIAM (2006) 5, 14
13. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision **40**(1) (2011) 120–145 7
14. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: International Conference on Computer Vision (ICCV 2011). (2011) 7
15. Alberti, G., Bouchitté, G., Maso, G.D.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. Calculus of Variations and Partial Differential Equations **16**(3) (2003) 299–333 7
16. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the piecewise smooth mumford-shah functional. In: IEEE Int. Conf. on Computer Vision, Kyoto, Japan (2009) 7, 8, 9
17. Rockafellar, R.T.: Convex Analysis. Princeton University Press (1996) 14

A   Appendix

Proof (Theorem 1). We have, using the divergence theorem,

$$
\begin{aligned}
R(u) &= \int_A (-\operatorname{div} p_i + q_i)\, dx \;+\; \int_{\bar{A}} (-\operatorname{div} p_j + q_j)\, dx \\
&= \int_{\partial A} (-p_i)\, \nu_{\partial A}\, d\mathcal{H}^1 + \int_{\partial \bar{A}} (-p_j)\, \nu_{\partial \bar{A}}\, d\mathcal{H}^1 + \int_A q_i\, dx + \int_{\bar{A}} q_j\, dx \\
&= \int_{\partial A} (p_j - p_i)\, \nu_{\partial A}\, d\mathcal{H}^1 + \int_A q_i\, dx + \int_{\bar{A}} q_j\, dx \qquad\qquad\text{(A.1)} \\
&\le \int_{\partial A} (p_j - p_i)\, \nu_{\partial A}\, d\mathcal{H}^1 + \int_A q_i\, dx.
\end{aligned}
$$

For the last inequality we used $q_j \le 0$, which follows from (2.3) by setting $i = j$. Using the constraints in (2.3), from this we obtain

$$
\begin{aligned}
R(u) &\le \int_{\partial A} (d(i,j) - q_i)\, d\mathcal{H}^1 + \int_A q_i\, dx \\
&= d(i,j)\operatorname{Per}(A) + \int_A q_i\, dx - \int_{\partial A} q_i\, d\mathcal{H}^1.
\end{aligned}
\qquad\text{(A.2)}
$$

Observe that in the discretized setting it holds

$$\int_{\partial A} q_i \, d\mathcal{H}^1 = \int_{A_1} q_i \, dx \qquad\qquad (A.3)$$

where $A_1 := \{x \in A \mid \text{dist}(x, \partial A) \leq 1\}$ are the points in A near its boundary. Hence, we have

$$\int_A q_i \, dx - \int_{\partial A} q_i \, d\mathcal{H}^1 = \int_{A \setminus A_1} q_i \, dx \leq 0 \qquad\qquad (A.4)$$

and it follows $R(u) \leq d(i,j) \, \text{Per}(A)$. It is also possible to show the equality here, i.e. that the supremum over p and q is reached. However, this requires a rather technical argument.

Proof (Theorem 2). The minimization of the overall functional (2.1) with the constraint $u \in K$ can be rewritten as unconstrained minimization of

$$E(u) = \sum_{i=1}^n \int_\Omega \varrho_i(x) \, u_i(x) \, dx \ + \ R(u) \ + \ \delta_K(u), \qquad\qquad (A.5)$$

over the space $V := L^2(\Omega; \mathbb{R}^n)$, where the indicator function $\delta_K(u)$ is defined as 0 for $u \in K$, and ∞ otherwise [17]. For $\varrho \in V$, the first term is a linear bounded function in u, and therefore convex and lower-semicontinuous on V. Also, δ_K is convex, coercive (since K is bounded), and lower-semicontinuous (since K is closed).

Together with Proposition 1 and $R(u) \geq 0$ (choose $(p,q) \equiv 0$ in (2.2)), we conclude that the overall energy E is convex, lower-semicontinuous and coercive on V. By [12, Theorem 3.3.3] E is then also weakly lower-semicontinuous and coercive. Existence of minimizers in V (which necessarily must lie in $K \subset V$ because of the delta function), now follows from [12, Theorem 3.2.5].

Proof (Proposition 2). Since the constant function u has zero gradient, the representation (2.2) of R reduces to

$$R(u) = \sup_{(p,q) \in C} \sum_{i=1}^n c_i \int_\Omega q_i(x) \, dx. \qquad\qquad (A.6)$$

with the set C in (2.3).

First, assume that $c_i \geq 0$ for all i. We can then "pull out" the constants c_i out of the supremum in (2.2). The constraints in C for $i = j$ specifically yield $q_i(x) \leq 0$ for all i and x. This way we get $R(u) \leq 0$. By choosing $p, q \equiv 0$ we also obtain $R(u) \geq 0$, so that overall $R(u) = 0$.

Assume now that $c_{i_0} < 0$ for some i_0. Choosing $p \equiv 0$, $q_i(x) := 0$ for $i \neq i_0$ and $q_{i_0}(x) := M$ with any $M \leq 0$, we obviously have $(p, q) \in C$. Therefore, $R(u) \geq \sup_{M \leq 0} c_{i_0} M |\Omega| = \infty$.