Generalized Roof Duality for Multi-Label Optimization: Optimal Lower Bounds and Persistency

Thomas Windheuser[1,2], Hiroshi Ishikawa[2], and Daniel Cremers[1]

[1] Department of Computer Science, Technische Universität München, Munich, Germany
[2] Department of Computer Science and Engineering, Waseda University, Tokyo, Japan

Abstract. We extend the concept of generalized roof duality from pseudo-boolean functions to real-valued functions over multi-label variables. In particular, we prove that an analogue of the persistency property holds for energies of any order with any number of linearly ordered labels. Moreover, we show how the optimal submodular relaxation can be constructed in the first-order case.

Key words: multi-label, higher-order, roof duality, MRF, computer vision

1 Introduction

Markov random fields are a standard optimization method for solving a variety of computer vision problems such as segmentation, denoising, stereo, and optical flow. While first-order binary submodular energies can be globally minimized [1], researchers have recently aimed at expanding their applicability to harder optimization problems that are

(NS) non-submodular,
(ML) multiple-label,
(HO) higher-order.

For instance, the swap and expansion algorithms [2] made it possible to approximately minimize non-submodular multi-label energies, thus tackling challenges (**NS**) and (**ML**). The QPBO algorithm allows optimization of many first-order non-submodular functions of binary variables at least partially - challenging (**NS**). Energies with linearly-ordered multiple labels can be globally minimized if convex (i.e., submodular) [3,4,5] (**ML**). Moreover, binary higher-order energies can be reduced [6,7,8,9,10] to first-order ones (**HO**).

In this paper, we propose to tackle the challenges of non-submodular multi-label functions (**NS**, **ML**) in a non-iterative way, with some theoretical results that apply also to higher-order energies (**HO**). We do this by extending the concept of roof duality [11,12,13,14], the basis of the QPBO algorithm, to energies with multiple labels. The concept was recently generalized to higher-order binary energies, i.e., to the direction of (**HO**), by Kolmogorov [15] and also by Kahl and Strandmark [9,10]. Here, we extend it toward multi-label (**ML**), using the non-iterative algorithms [3,4,5]. Similarly to the binary case, we construct the tightest lower-bound submodular function of the given energy and minimize the latter instead of the original. The key contribution of this work

original noisy lower bounds upper bounds

Fig. 1. Image denoising as an example application of the generalized roof dual
The task of removing noise from an image can be modelled as the problem of optimizing the (generally NP-hard) multi-label functional (20). (Original image on the left and gaussian noise with $\sigma = 0.2$ added to the second image from the left.) We show that (given a fixed ordering of the labels) the optimal submodular relaxation that yields the maximum lower bound of the functional can be computed in polynomial time. This maximum lower bound is called the generalized roof dual. Due to the Persistency Theorem 1, the computed optimal solution to the submodular relaxation guarantees, for each pixel, the intensity value of an optimal solution of the original optimization problem to be in a certain range. The image second from the right shows the lower bounds of these ranges and the rightmost image shows the upper bounds. On average, only 1.2 intensity values out of 40 per pixel and color channel are possible candidates for the optimal solution. Thus each of the solutions to the relaxed problem is very close to the optimal solution of the original problem.

is to prove that an analogue of the persistency property also holds for energies of any order with any number of linearly ordered labels.

Moreover, we show how the optimal submodular relaxations can be constructed in the first-order case. This construction turns out to coincide with the algorithm of Kohli et al. [16], which tries to tackle first-order multi-label problems and also has the persistency property. In this regard we show that the algorithm of [16] actually computes the optimal submodular relaxation in the sense of generalized roof duality.

The meaning of Theorem 1, the main result of this paper that proves the persistency property, is illustrated in Figure 1. The persistency gives us information about an optimal labeling in the form of the range of labels at each pixel in which an optimal label must lie. In the case of binary QPBO, we acquire no information at all if a pixel is not labeled. In contrast, according to the multi-label generalization presented in this paper, we can often exclude labels that are definitely not part of any optimal solution even if we do not get a partial labeling, effectively reducing the search space of the optimization problem.

The paper is structured as follows. In the next section, we fix notation. In section 3, we define the generalized roof dual for multi-label energies. In section 4, we discuss the main theorem of this paper, which says that a persistency property holds for the generalized roof dual for multi-label energies of any order. In section 5, we give a closed-formula optimum lower bound in the case of first order energies. In section 6, the exact construction of the optimal submodular relaxation of given general first-order multi-label function is explained. The experimental section 7 investigates our approach of computing the generalized roof dual in an image denoising example. To improve readability we deferred all proves to the appendix.

2 Notation

Throughout this paper, we deal with variables that can take values in a finite label set with a linear order. Thus, whenever we use such a label set $\mathcal{L}$, we identify it with $\mathcal{L} := \{0, \ldots, \ell - 1\}$. We define the negation of a scalar multi-labeled variable $x \in \mathcal{L}$ analogous to its boolean counterpart by

$$\overline{x} := \ell - (x + 1).$$

For vectors $\mathbf{x} \in \mathcal{L}^n$ we extend the operators $\overline{\cdot}, +, -, \min$, and $\max$ component-wise.

3 Generalizing Roof Duality to Multi-Label Problems

The goal of multi-label minimization is to minimize a real-valued function $f : \mathcal{L}^n \to \mathbb{R}$ of n variables that can take values from a finite set $\mathcal{L}$, called the label set:

$$\min_{\mathbf{x} \in \mathcal{L}^n} f(\mathbf{x}).$$

Of course, this problem is at least as hard as pseudo-boolean optimization (which is the special case that the set $\mathcal{L}$ contains only two values). It is easier to minimize the following submodular function that gives a lower bound of f:

$$\min_{(\mathbf{x},\mathbf{y}) \in \mathcal{L}^{2n}} g(\mathbf{x}, \mathbf{y}),$$

where $g : \mathcal{L}^{2n} \to \mathbb{R}$ is such that

$$\forall \mathbf{x} \in \mathcal{L}^n : g(\mathbf{x}, \overline{\mathbf{x}}) = f(\mathbf{x}), \tag{1}$$

$$g \text{ is submodular, and} \tag{2}$$

$$\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{L}^{2n} : g(\mathbf{x}, \mathbf{y}) = g(\overline{\mathbf{y}}, \overline{\mathbf{x}}) \text{ (symmetry).} \tag{3}$$

Property (1) assures that the image of f is included in the image of g. As a consequence,

$$\min_{(\mathbf{x},\mathbf{y}) \in \mathcal{L}^{2n}} g(\mathbf{x}, \mathbf{y}) \le \min_{\mathbf{x} \in \mathcal{L}^n} f(\mathbf{x})$$

and the minimum of g is a lower bound of f. The symmetry property (3) allows to prove a persistency statement, which allows to obtain a partial labeling of f from a minimizer of g and will be explained later. The central property that makes g easier to solve than f is the submodularity. Originally defined for pseudo-boolean functions, it can be extended to linearly-ordered multi-labeled functions as well.

Definition 1 **([17]).** *Given a linearly-ordered finite set* $\mathcal{L} = \{0, \ldots, \ell - 1\}$*, a function* $\theta : \mathcal{L} \times \mathcal{L} \to \mathbb{R}$ *is said to be* submodular *if for any four labels* $x_1, x_2, y_1, y_2 \in \mathcal{L}$ *that satisfy* $x_1 < x_2, y_1 < y_2$*, the inequality*

$$\theta(x_1, y_1) + \theta(x_2, y_2) \leq \theta(x_1, y_2) + \theta(x_2, y_1)$$

holds.

Definition 2 **([17]).** *Given a function* $g : \mathcal{L}^n \to \mathbb{R}$ *and* $\mathbf{z} \in \mathcal{L}^n$*, the* projection *of* g *to its* i*-th and* j*-th components* $(i, j \in \{1, \ldots, n\}, i < j)$ *is*

$$g_{[i,j,\mathbf{z}]}(x, y) := g(z_1, \ldots, z_{i-1}, x, z_{i+1}, \ldots, z_{j-1}, y, z_{j+1} \ldots, z_n).$$

Then the function g *is said to be* submodular *if for all* $i, j \in \{1, \ldots, n\}, i < j$ *and all* $\mathbf{z} \in \mathcal{L}^n$ *the projection* $g_{[i,j,\mathbf{z}]}$ *is submodular.*

This definition of submodularity is due to Flach and Schlesinger [17], who denominate it as subconvexity. Throughout this paper will will use the term submodularity to emphasize the connection to pseudo-boolean optimization.

For a given f, there are infinitely many choices for g in general. Since g is a lower bound, one is interested in finding the g with the highest minimum and thus the highest lower bound for f.

$$\begin{aligned} &\max_{g} \min_{(\mathbf{x},\mathbf{y}) \in \mathcal{L}^{2n}} g(\mathbf{x}, \mathbf{y}) \\ &\text{s. t. } \; g \text{ satisfies (1)-(3).} \end{aligned} \tag{4}$$

We call the maximum of (4) the *generalized roof dual* of f. This is clearly a generalization of the generalized roof dual by Kahl and Strandmark [9,10]. In this paper, we will show that their persistency guarantees still hold even if we move from functions on boolean variables to functions on multi-labeled variables.

It would be beneficial if the whole optimization problem could be stated as a linear program. We will see later that the requirements (1),(3) (and with some additional assumptions (2)) can be expressed by linear constraints. The (potentially) higher-order objective function g can also be transformed into linear constraints by introducing a variable l for the minimum of g:

$$\begin{aligned} &\max_{g,l} \; l \\ &\text{s. t. } \; \forall (\mathbf{x}, \mathbf{y}) \in \mathcal{L}^{2n} : g(\mathbf{x}, \mathbf{y}) \geq l \\ &\qquad g \text{ satisfies (1)-(3).} \end{aligned} \tag{5}$$

Keep in mind that the number of constraints including l can be exponential in n.

We end this section by stating that the symmetry constraint does not affect the maximum of (5).

Lemma 1. *Given a function* $f : \mathcal{L}^n \to \mathbb{R}$*, there exists a* $g^* : \mathcal{L}^{2n} \to \mathbb{R}$ *defined by*

$$\begin{aligned} (g^*, l^*) \in \arg\max_{g,l} \; & l \\ \text{s. t. } \; & \forall (\mathbf{x}, \mathbf{y}) \in \mathcal{L}^{2n} : g(\mathbf{x}, \mathbf{y}) \geq l \\ & \forall \mathbf{x} \in \mathcal{L}^n : g(\mathbf{x}, \overline{\mathbf{x}}) = f(\mathbf{x}), \\ & g \text{ is submodular,} \end{aligned} \tag{6}$$

such that g^* *is symmetric, i.e.* $\forall \mathbf{x}, \mathbf{y} \in \mathcal{L}^n : g^*(\mathbf{x}, \mathbf{y}) = g^*(\overline{\mathbf{y}}, \overline{\mathbf{x}})$.

4 Persistency

In this section, we show that a global minimizer of any submodular relaxation g of f defines a range in which a global minimum of f must lie. Moreover moving any point into this range does not increase the energy with respect to f. This property is called *persistency*.

We begin by defining the subset $\mathcal{S}^n \subset \mathcal{L}^{2n}$ as

$$\mathcal{S}^n := \{(\mathbf{x}, \mathbf{y}) \in \mathcal{L}^{2n} \mid \forall i \in \{1, \ldots, n\} : x_i + y_i < \ell\}.$$

The following lemma tells us that there is alway a minimizer of g lying in $\mathcal{S}^n$:

Lemma 2. *For any submodular symmetric function* $g : \mathcal{L}^{2n} \to \mathbb{R}$ *the following statement is true:*

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{L}^n : g(\mathbf{x}, \mathbf{y}) \geq g(\min(\mathbf{x}, \overline{\mathbf{y}}), \min(\overline{\mathbf{x}}, \mathbf{y})). \tag{7}$$

Since $(\min(\mathbf{x}, \overline{\mathbf{y}}), \min(\overline{\mathbf{x}}, \mathbf{y})) \in \mathcal{S}^n$*, there always exists a point in* $\mathcal{S}^n$ *that minimizes* g.

Equation (7) allows us to transform any minimizer of g into an element of $\mathcal{S}^n$ that still minimizes g. Interestingly a point $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}^n$ defines a range of labels for each variable, since for any i the inequalities $1 \leq x_i \leq \overline{y}_i \leq \mathcal{L}$ are true. The next definition fixes the projection of a point in $\mathcal{L}^n$ onto these ranges.

Definition 3. *For any point* $\mathbf{x} \in \mathcal{L}^n$ *and* $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{S}^n$*, the* overwrite operator $\mathcal{L}^n \times \mathcal{S}^n \to \mathcal{L}^n$ *denoted by* $\mathbf{x} \leftarrow (\mathbf{x}^*, \mathbf{y}^*)$ *is defined component-wise by*

$$[\mathbf{x} \leftarrow (\mathbf{x}^*, \mathbf{y}^*)]_i = \begin{cases} x_i^* & \text{if } x_i < x_i^*, \\ \overline{y}_i^* & \text{if } x_i > \overline{y}_i^*, \\ x_i & \text{otherwise.} \end{cases}$$

Now we come to the central persistency statement that tells us that parts of the minimizer of g can be optimal for function f. We can overwrite any point $\mathbf{x} \in \mathcal{L}^n$ by these parts and still get a lower or equal value for f.

Theorem 1 (Persistency). *Let* g *be a function satisfying* (1)-(3) *and* $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{S}^n$ *be a minimizer of* g*, then* $\forall \mathbf{x} \in \mathcal{L}^n : f(\mathbf{x} \leftarrow (\mathbf{x}^*, \mathbf{y}^*)) \leq f(\mathbf{x})$. *In particular, if* $\mathbf{x} \in \arg\min(f)$, *then also* $\mathbf{x} \leftarrow (\mathbf{x}^*, \mathbf{y}^*) \in \arg\min(f)$.

Theorem 1 tells us that, similarly to binary labeling, if $x_i^* = \overline{y}_i^*$ then setting $x_i = x_i^*$ is an optimal choice. It does not increase the energy, no matter what the other values of $\mathbf{x}$ are. What is different in the multi-label case is that a label pair x_i^*, y_i^* contains information about an optimal labeling of f even if $x_i^* < \overline{y}_i^*$. Combining Theorem 1 and Definition 3 for any optimizer $\mathbf{x}$ of f, we can assume without loss of generality that the value of variable x_i satisfies $x_i^* \leq x_i \leq \overline{y}_i^*$. If x_i lies outside these boundaries we can move it into the inside without losing the optimality of $\mathbf{x}$. Thus, labels outside the range $\{x_i^*, \ldots, \overline{y}_i^*\}$ don't have to be taken into account when optimizing f. This effectively reduces the search space of the optimization problem. Only in the case where $(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{0}, \mathbf{0})$ we do not acquire any information at all.

5 First Order Case

We show that if f has degree 2 the optimum g of (4) can be expressed by a closed formula. First, note that any degree 2 function can be written in the form

$$f(\mathbf{x}) = \sum_{i<j} \gamma_{ij}(x_i, x_j), \tag{8}$$

by adding the terms of the form $\gamma_{ji}(x_j, x_i)$ with $i < j$ to $\gamma_{ij}(x_i, x_j)$ and those of the form $\gamma_i(x_i)$ and $\gamma_i(x_i, x_i)$ to one of the terms of the form $\gamma_{ij}(x_i, x_j)$ or $\gamma_{ki}(x_k, x_i)$.

We begin by fixing a representation of g that directly includes the symmetry constraint (3).

Lemma 3. *Any symmetric function $g : \mathcal{L}^{2n} \to \mathbb{R}$ of degree 2 can be represented as*

$$g(\mathbf{x}, \mathbf{y}) := \frac{1}{2} \sum_{i \le j} \left(\theta_{ij}(x_i, x_j) + \theta_{ij}(\overline{y}_i, \overline{y}_j) + \theta'_{ij}(x_i, \overline{y}_j) + \theta'_{ij}(\overline{y}_i, x_j)\right) \tag{9}$$

using a set of functions $\theta := \{\theta_{ij}, \theta'_{ij}\}_{i \le j}$, which we call a parameter set *of g.*

Let f be given in the form (8) (we assume $\gamma_{ii}(x, y) = 0$) and compare it with

$$g(\mathbf{x}, \overline{\mathbf{x}}) := \sum_{i \le j} \left(\theta_{ij}(x_i, x_j) + \theta'_{ij}(x_i, x_j)\right).$$

Lemma 4. *If g satisfies $f(\mathbf{x}) = g(\mathbf{x}, \overline{\mathbf{x}})$, there is a parameter set θ of g that satisfies*

$$\gamma_{ij}(x, y) = \theta_{ij}(x, y) + \theta'_{ij}(x, y) \text{ for all } i \le j. \tag{10}$$

Let us denote

$$\alpha_{ij}(x_1, x_2, y_1, y_2) := \gamma_{ij}(x_1, y_1) + \gamma_{ij}(x_2, y_2) - \gamma_{ij}(x_1, y_2) - \gamma_{ij}(x_2, y_1),$$
$$\alpha^-_{ij}(x_1, x_2, y_1, y_2) := \min\left(0, \alpha_{ij}(x_1, x_2, y_1, y_2)\right),$$

and for a parameter set θ of g,

$$\beta^\theta_{ij}(x_1, x_2, y_1, y_2) := \theta_{ij}(x_1, y_1) + \theta_{ij}(x_2, y_2) - \theta_{ij}(x_1, y_2) - \theta_{ij}(x_2, y_1), \text{ and}$$
$$\beta'^\theta_{ij}(x_1, x_2, y_1, y_2) := \theta'_{ij}(x_1, y_1) + \theta'_{ij}(x_2, y_2) - \theta'_{ij}(x_1, y_2) - \theta'_{ij}(x_2, y_1).$$

Now we can rewrite our optimization problem (4) as

$$\max_{\theta} \min_{\mathbf{x}, \mathbf{y} \in \mathcal{L}^{2n}} g(\mathbf{x}, \mathbf{y}) \tag{11}$$
$$= \frac{1}{2} \sum_{i \le j} \left(\theta_{ij}(x_i, x_j) + \theta_{ij}(\overline{y}_i, \overline{y}_j) + \theta'_{ij}(x_i, \overline{y}_j) + \theta'_{ij}(\overline{y}_i, x_j)\right)$$

$$\text{s. t. } \forall i \le j : \forall x, y \in \mathcal{L} : \quad \gamma_{ij}(x, y) = \theta_{ij}(x, y) + \theta'_{ij}(x, y), \tag{12}$$

$$\forall i \le j : \forall x_1, x_2, y_1, y_2 \in \mathcal{L}, x_1 \prec x_2, y_1 \prec y_2 : \quad 0 \ge \beta^\theta_{ij}(x_1, x_2, y_1, y_2), \tag{13}$$

$$\forall i \le j : \forall x_1, x_2, y_1, y_2 \in \mathcal{L}, x_1 \prec x_2, y_1 \prec y_2 : \quad 0 \le \beta'^\theta_{ij}(x_1, x_2, y_1, y_2), \tag{14}$$

Note that because of (12) and $\gamma_{ii}(x, y) = 0$, $\theta_{ii}(x, y) = -\theta'_{ii}(x, y)$ and therefore for $i = j$ (13) is automatically satisfied if (14) is, and vice versa.

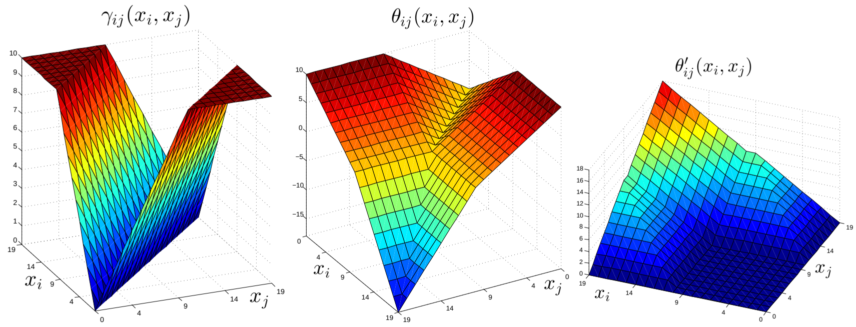

Fig. 2. Optimal submodular relaxation of a pairwise term γ_{ij}.
The plots of of a non-submodular function $\gamma_{ij}(x_i, x_j) = \min(|x_i - x_j|, 10)$ and its optimal submodular relaxation $\theta_{ij}, \theta'_{ij}$. They are derived from γ_{ij} by the rules explained in Section 6. Note that while θ_{ij} is submodular, the function $\theta'_{ij} = \gamma_{ij} - \theta_{ij}$ is only submodular when one of its arguments is flipped.

Theorem 2. *Let θ^* be a parameter set of g satisfying (12) and*

$$\beta_{ij}^{\theta^*}(x_1, x_2, y_1, y_2) = \sum_{x_1 \le k < x_2} \sum_{y_1 \le l < y_2} \alpha_{ij}^{-}(l, l+1, k, k+1). \tag{15}$$

Then θ^* is a maximizer of (11).

6 Implementation

In this section, we will specify how to find an optimal submodular relaxation g of a given general first-order multi-label function f. Consider f having the form of $f(\mathbf{x}) = \sum_{i \le j} \gamma_{ij}(x_i, x_j)$. The optimal relaxations $\theta_{ij}, \theta'_{ij}$ for each pairwise term can be construct by setting

$$\begin{aligned}
&\forall i \le j : \forall x \in \mathcal{L} : \\
&\qquad \theta_{ij}(0, x) = \gamma_{ij}(x, 0), && (16)\\
&\qquad \theta_{ij}(x, 0) = \gamma_{ij}(0, x), && (17)\\
&\forall i \le j : \forall x_i, x_j \in \mathcal{L}, x_i > 0, x_j > 0 : \\
&\quad \theta_{ij}(x_i, x_j) = \alpha_{ij}^{-}(x_i, x_i - 1, x_j, x_j - 1) \\
&\qquad\qquad - \theta_{ij}(x_i - 1, x_j - 1) + \theta_{ij}(x_i, x_j - 1) + \theta_{ij}(x_i - 1, x_j) \text{ and} && (18)\\
&\forall i \le j : \forall x_i, x_j \in \mathcal{L} : \\
&\quad \theta'_{ij}(x_i, x_j) = \gamma_{ij}(x_i, x_j) - \theta_{ij}(x_i, x_j). && (19)
\end{aligned}$$

Where the terms $\theta_{ij}(x_i, x_j)$ in (18) are defined iteratively by increasing labels x_i and x_j. It can be easily verified that this parameter set satisfies (12) and (15). Due to Theorem

2 this parameter set defines a submodular function g that attains the maximum (4) and thus the generalized roof dual. The submodular function g can be optimized globally in polynomial time by finding a minimum cut in a special graph using the construction of Schlesinger and Flach [4].

We note that this construction of g coincides with the construction by Kohli et al. [16] after the application of QPBO. Thus by Theorem 2 we have shown the optimality of [16] regarding the generalized roof dual for multi-label problems.

In Figure 2 we show an example of an optimal submodular relaxation of a non-submodular function $\gamma_{ij}(x_i, x_j) = \min(|x_i - x_j|, 10)$.

7 Image Denoising Example

The goal of image denoising is to remove the noise from a noisy image. In order to cast the denoising problem as a multi-label MRF, we consider a grey-scale image as a discrete function $\mathcal{I} : \mathcal{P} \to \{0, .., 255\}$ defined on a finite set of pixels $\mathcal{P}$. On the one hand, the denoised image $\mathbf{x}^* : \mathcal{P} \to \mathcal{L} \subset \{0, .., 255\}$ should be close to the original image $\mathcal{I}$; but on the other hand, it should be less noisy. We define the denoised image as the minimum of the first-order multi-label MRF energy

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_{p \in \mathcal{P}} D_p(\mathbf{x}) + \lambda \sum_{pq \in \mathcal{N}} S_{pq}(\mathbf{x}), \tag{20}$$

where $\mathcal{N}$ is a neighborhood structure on the pixels, $\lambda \in \mathbb{R}_{\geq 0}$ is a parameter, and D_p, S_{pq} are the data terms and smoothness terms, respectively. In all our experiments, we used the 4-neighborhood as the neighborhood structure. The data term $D_p(\mathbf{x})$ makes $\mathbf{x}$ to stay close to $\mathcal{I}$ and is defined as $D_p(\mathbf{x}) = (\mathcal{I}(p) - \mathbf{x}(p))^2$. We assume that a denoised image is much smoother than the given noisy image. The so-called smoothness prior penalizes strong gradients in $\mathbf{x}$ by measuring the absolute differences of values between two neighboring pixels truncated by some maximum value $T \in \mathbb{N}$. It is defined as $S_{pq}(\mathbf{x}) = \min(|\mathbf{x}(p) - \mathbf{x}(q)|, T)$. This pairwise term is clearly non-submodular. Its plot and that of its submodular relaxation are shown in Figure 2.

By constructing the optimal submodular relaxation g, we can compute the maximum lower bound of the optimal $\mathbf{x}^*$ with respect to (4). Moreover, the minimizer of g also yields a partial labeling. A partial labeling for some variable indicates constraints on the labels that can be assigned to this variable. In the case of image denoising, these constraints are lower and upper bounds on the intensity values of each pixel. In Figure 3, the denoising of a small image with different parameters is shown ($|\mathcal{L}| = 32$ and the standard deviation of the added gaussian noise $\sigma = 0.2$). On the right, the upper and lower bounds for three different parameter values are shown. The Persistency Theorem 1 guarantees that the optimal denoised image is between these two bounds. For parameters $T = 6$ and $T = 8$, the range in which the optimum lies is rather large; on average, the widths of the range from which each pixel is allowed to take a value are 17.60 and 17.03, respectively. For parameter $T = 10$, on average the range in which the optimal denoised pixel lies is 1.09. On the left of Figure 3, a plot illustrates the connection between different values for parameters T and λ and the size of the range in which the optimal denoised pixels lie.

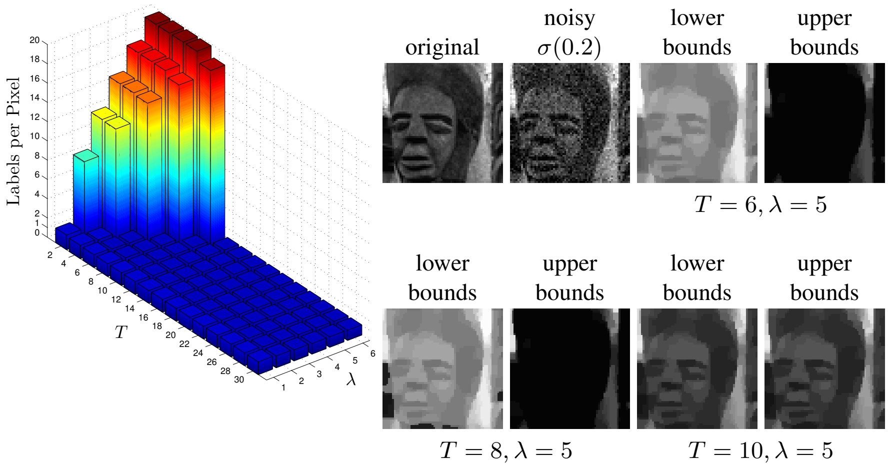

Fig. 3. Computing the generalized roof dual for an image denoising functional (20) with different parameters.

Noise is removed from an image by using energy functional (20) with different parameter values. Due to the Persistency Theorem 1, the computed minimizer of the submodular relaxation yields for each pixel a range in which the optimal labeling must lie. The plot on the left shows the average size of this range per pixel for different parameter values. The smaller the range, the closer the solution of the submodular relaxation is to the optimal solution of the original problem. On the right, the images show the lower and upper bounds of these ranges for three parameter values. For $T = 6$ and $T = 8$, the average sizes are 17.60 and 17.03, respectively, whereas for $T = 10$ the average size is only 1.09.

8 Conclusion

We generalized the concept of roof duality from binary labeling problems to multi-label problems where labels exhibit a linear ordering. As a consequence, we can tackle non-submodular multi-label problems in a direct and non-iterative manner. In particular, we prove that an analogue of the persistency property prevails in the multi-label scenario of arbitrary order, which gives us information about an optimal labeling by allowing the exclusion of labels at each pixel that are certainly not part of any optimal solution. Moreover, we show how the optimal submodular relaxation can be constructed for non-submodular first-order case of multi-label problems.

Acknowledgment This work was partially supported by the Erasmus Mundus-BEAM program, the JST Mathematics Program Sakigake (PRESTO), the JSPS Grant-in-Aid for Scientific Research (B) #24300075, the MEXT Grant-in-Aid for Scientific Research on Innovative Areas #24103709, and the ERC Starting grant ‘ConvexVision’.

A Proof of Symmetry

Proof (Lemma 1). Given a function $g : \mathcal{L}^{2n} \to \mathbb{R}$ we define its symmetric and asymmetric parts by

$$g_{\text{sym}}(\mathbf{x}, \mathbf{y}) := 1/2(g(\mathbf{x}, \mathbf{y}) + g(\overline{\mathbf{y}}, \overline{\mathbf{x}})),$$
$$g_{\text{asym}}(\mathbf{x}, \mathbf{y}) := 1/2(g(\mathbf{x}, \mathbf{y}) - g(\overline{\mathbf{y}}, \overline{\mathbf{x}})).$$

We see that the decomposition

$$g(\mathbf{x}, \mathbf{y}) = g_{\text{sym}}(\mathbf{x}, \mathbf{y}) + g_{\text{asym}}(\mathbf{x}, \mathbf{y})$$

follows from the definitions. We show that if (g, l) is a maximum of (6), then (g_{sym}, l) is also a maximum.

To see $g_{\text{sym}}(\mathbf{x}, \mathbf{y}) \geq l$ for any point $(\mathbf{x}, \mathbf{y})$, using the maximality of (g, l), we have

$$g(\mathbf{x}, \mathbf{y}) = g_{\text{sym}}(\mathbf{x}, \mathbf{y}) + g_{\text{asym}}(\mathbf{x}, \mathbf{y}) \geq l \text{ and}$$
$$g(\overline{\mathbf{y}}, \overline{\mathbf{x}}) = g_{\text{sym}}(\mathbf{x}, \mathbf{y}) - g_{\text{asym}}(\mathbf{x}, \mathbf{y}) \geq l \ .$$

Adding both inequalities we get the desired result

$$2g_{\text{sym}}(\mathbf{x}, \mathbf{y}) \geq 2l \ .$$

The condition $g_{sym}(\mathbf{x}, \overline{\mathbf{x}}) = f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{L}^n$ follows from

$$\forall \mathbf{x} \in \mathcal{L}^n : g_{\text{sym}}(\mathbf{x}, \overline{\mathbf{x}}) = 1/2(g(\mathbf{x}, \overline{\mathbf{x}}) + g(\overline{\overline{\mathbf{x}}}, \overline{\mathbf{x}})) = g(\mathbf{x}, \overline{\mathbf{x}}) = f(\mathbf{x}).$$

It remains to show submodularity of g_{sym}. Choose and fix $i, j \in \{1, \ldots, 2n\}, i < j$ and $(\mathbf{z}, \mathbf{z}') \in \mathcal{L}^{2n}$. Let $i' := (i + n) \bmod 2n$ and $j' := (j + n) \bmod 2n$. By definition, the projection of g_{sym} to $i, j, (\mathbf{z}, \mathbf{z}')$ can be expressed as

$$g_{\text{sym}[i,j,(\mathbf{z},\mathbf{z}')]}(x, y) := 1/2 \left(g_{[i,j,(\mathbf{z},\mathbf{z}')]}(x, y) + g_{[i',j',(\overline{\mathbf{z}'},\overline{\mathbf{z}})]}(\overline{x}, \overline{y}) \right) .$$

Because g is submodular, for any $x_1, x_2, y_1, y_2 \in \mathcal{L}$ satisfying $x_1 < x_2, y_1 < y_2$, the two inequalities

$$g_{[i,j,(\mathbf{z},\mathbf{z}')]}(x_1, y_1) + g_{[i,j,(\mathbf{z},\mathbf{z}')]}(x_2, y_2) \leq g_{[i,j,(\mathbf{z},\mathbf{z}')]}(x_2, y_1) + g_{[i,j,(\mathbf{z},\mathbf{z}')]}(x_1, y_2)$$
$$g_{[i',j',(\overline{\mathbf{z}'},\overline{\mathbf{z}})]}(\overline{x_1}, \overline{y_1}) + g_{[i',j',(\overline{\mathbf{z}'},\overline{\mathbf{z}})]}(\overline{x_2}, \overline{y_2}) \leq g_{[i',j',(\overline{\mathbf{z}'},\overline{\mathbf{z}})]}(\overline{x_2}, \overline{y_1}) + g_{[i',j',(\overline{\mathbf{z}'},\overline{\mathbf{z}})]}(\overline{x_1}, \overline{y_2})$$

hold. Adding both inequalities yields the submodularity condition for g_{sym}, concluding the proof. □

B Proof of Persistency

The proof of the persistency statement needs several preliminary observations. The first one is a equivalent definition of submodularity.

Lemma 5 ([17] Lemma 1). *$g : \mathcal{L}^n \to \mathbb{R}$ is submodular if and only if*

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{L}^n : g(\mathbf{x}) + g(\mathbf{y}) \geq g(\min(\mathbf{x}, \mathbf{y})) + g(\max(\mathbf{x}, \mathbf{y})).$$

This allows us to prove Lemma 2.

Proof (Lemma 2). For any point $(\mathbf{x}, \mathbf{y}) \in \mathcal{L}^{2n}$ it holds:

$$\begin{aligned} g(\mathbf{x}, \mathbf{y}) &= 1/2(g(\mathbf{x}, \mathbf{y}) + g(\overline{\mathbf{y}}, \overline{\mathbf{x}})) && \text{(symmetry)} \\ &\geq 1/2(g(\min(\mathbf{x}, \overline{\mathbf{y}}), \min(\overline{\mathbf{x}}, \mathbf{y})) + g(\max(\mathbf{x}, \overline{\mathbf{y}}), \max(\overline{\mathbf{x}}, \mathbf{y}))) && \text{(Lemma (5))} \\ &= 1/2(g(\min(\mathbf{x}, \overline{\mathbf{y}}), \min(\overline{\mathbf{x}}, \mathbf{y})) + g(\overline{\min(\overline{\mathbf{x}}, \mathbf{y})}, \overline{\min(\mathbf{x}, \overline{\mathbf{y}})})) && (\max(x, y) = \overline{\min(\overline{x}, \overline{y})}) \\ &= g(\min(\mathbf{x}, \overline{\mathbf{y}}), \min(\overline{\mathbf{x}}, \mathbf{y})). && \text{(symmetry)} \end{aligned}$$

Finally observe that $(\min(\mathbf{x}, \overline{\mathbf{y}}), \min(\overline{\mathbf{x}}, \mathbf{y})) \in \mathcal{S}^n$. □

The next lemma is directly used in the proof of persistency.

Lemma 6. *Let $g : \mathcal{S}^n \to \mathbb{R}$ be a submodular, symmetric function, for any point $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}^n$ and any minimizer $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{S}^n$ of g the following inequality is true:*

$$g(\mathbf{x}, \mathbf{y}) \geq g(\min(\max(\mathbf{x}, \mathbf{x}^*), \overline{\mathbf{y}}, \overline{\mathbf{y}}^*), \min(\max(\mathbf{y}, \mathbf{y}^*), \overline{\mathbf{x}}, \overline{\mathbf{x}}^*)).$$

Proof. From submodularity and Lemma 5 we have

$$g(\mathbf{x}, \mathbf{y}) + g(\mathbf{x}^*, \mathbf{y}^*) \geq g(\min(\mathbf{x}, \mathbf{x}^*), \min(\mathbf{y}, \mathbf{y}^*)) + g(\max(\mathbf{x}, \mathbf{x}^*), \max(\mathbf{y}, \mathbf{y}^*)). \quad (21)$$

Because $(\mathbf{x}^*, \mathbf{y}^*)$ is a minimizer of g it is also true that

$$g(\mathbf{x}^*, \mathbf{y}^*) \leq g(\min(\mathbf{x}, \mathbf{x}^*), \min(\mathbf{y}, \mathbf{y}^*)). \quad (22)$$

Using (22) to simplify (21) and applying Lemma 2 afterwards yields the proposition:

$$\begin{aligned} g(\mathbf{x}, \mathbf{y}) &\geq g(\max(\mathbf{x}, \mathbf{x}^*), \max(\mathbf{y}, \mathbf{y}^*)) \\ &\geq g(\min(\max(\mathbf{x}, \mathbf{x}^*), \overline{\max(\mathbf{y}, \mathbf{y}^*)}), \min(\overline{\max(\mathbf{x}, \mathbf{x}^*)}, \max(\mathbf{y}, \mathbf{y}^*))) \\ &= g(\min(\max(\mathbf{x}, \mathbf{x}^*), \overline{\mathbf{y}}, \overline{\mathbf{y}}^*), \min(\overline{\mathbf{x}}, \overline{\mathbf{x}}^*, \max(\mathbf{y}, \mathbf{y}^*))) \end{aligned}$$

□

Before we come to the final proof we state a direct consequence of the overwrite operator's definition:

Corollary 1. *For any point $\mathbf{x} \in \mathcal{L}^n$ and $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{S}^n$ define $\mathbf{z} := \mathbf{x} \leftarrow (\mathbf{x}^*, \mathbf{y}^*)$. It holds*

$$\begin{aligned} z_i &= \min(\overline{y}_i^*, \max(x_i^*, x_i)) \text{ and} \\ \overline{z}_i &= \min(\overline{x}_i^*, \max(y_i^*, \overline{x}_i)). \end{aligned}$$

Proof (Theorem 1). For any $\mathbf{x} \in \mathcal{L}^n$ it holds

$$
\begin{aligned}
f(\mathbf{x}) &= g(\mathbf{x}, \overline{\mathbf{x}}) && \text{(Property (1))}\\
&\geq g(\min(\max(\mathbf{x}, \mathbf{x}^*), \mathbf{x}, \overline{\mathbf{y}}^*), \min(\overline{\mathbf{x}}, \overline{\mathbf{x}}^*, \max(\overline{\mathbf{x}}, \mathbf{y}^*))) && \text{(Lemma 6)}\\
&= g(\min(\mathbf{x}, \overline{\mathbf{y}}^*), \min(\overline{\mathbf{x}}, \overline{\mathbf{x}}^*)) && \text{(simplification)}\\
&\geq \begin{array}{l} g(\ \min(\max(\min(\mathbf{x}, \overline{\mathbf{y}}^*), \mathbf{x}^*), \overline{\min(\overline{\mathbf{x}}, \overline{\mathbf{x}}^*)}, \overline{\mathbf{y}}^*),\\ \quad \min(\max(\min(\overline{\mathbf{x}}, \overline{\mathbf{x}}^*), \mathbf{y}^*), \overline{\min(\mathbf{x}, \overline{\mathbf{y}}^*)}, \overline{\mathbf{x}}^*)) \end{array} && \text{(Lemma 6)}\\
&= \begin{array}{l} g(\ \min(\max(\min(\mathbf{x}, \overline{\mathbf{y}}^*), \mathbf{x}^*), \max(\mathbf{x}, \mathbf{x}^*), \overline{\mathbf{y}}^*),\\ \quad \min(\max(\min(\overline{\mathbf{x}}, \overline{\mathbf{x}}^*), \mathbf{y}^*), \max(\overline{\mathbf{x}}, \mathbf{y}^*), \overline{\mathbf{x}}^*)) \end{array} && \text{(simplification)}\\
&= g(\min(\max(\mathbf{x}, \mathbf{x}^*), \overline{\mathbf{y}}^*), \min(\max(\overline{\mathbf{x}}, \mathbf{y}^*), \overline{\mathbf{x}}^*)) && \text{(simplification)}\\
&= g(\mathbf{x} \leftarrow (\mathbf{x}^*, \mathbf{y}^*), \overline{\mathbf{x} \leftarrow (\mathbf{x}^*, \mathbf{y}^*)}). && \text{(Corollary 1)}
\end{aligned}
$$

□

C Proofs of the First Order Case

Proof (Lemma 3). Any degree 2 function $g(\mathbf{x}, \mathbf{y})$ can be written

$$g(\mathbf{x}, \mathbf{y}) = \sum_{i \leq j} \left(\chi_{ij}(x_i, x_j) + \chi'_{ij}(\overline{y}_i, \overline{y}_j) + \chi''_{ij}(x_i, \overline{y}_j) + \chi'''_{ij}(\overline{y}_i, x_j)\right).$$

Then we have

$$
\begin{aligned}
g(\mathbf{x}, \mathbf{y}) + g(\overline{\mathbf{y}}, \overline{\mathbf{x}}) = \sum_{i \leq j} \big(&\chi_{ij}(x_i, x_j) + \chi'_{ij}(x_i, x_j) + \chi_{ij}(\overline{y}_i, \overline{y}_j) + \chi'_{ij}(\overline{y}_i, \overline{y}_j) + \\
&\chi''_{ij}(x_i, \overline{y}_j) + \chi'''_{ij}(x_i, \overline{y}_j) + \chi''_{ij}(\overline{y}_i, x_j) + \chi'''_{ij}(\overline{y}_i, x_j)\big).
\end{aligned}
$$

The proposition follows if we define

$$\theta_{ij}(x, y) = \chi_{ij}(x, y) + \chi'_{ij}(x, y), \quad \theta'_{ij}(x, y) = \chi''_{ij}(x, y) + \chi'''_{ij}(x, y).$$

□

Proof (Lemma 4). Let $\bar{\theta}$ be a parameter set of g. Define

$$\eta_{ij}(x, y) = \bar{\theta}_{ij}(x, y) + \bar{\theta}'_{ij}(x, y) - \gamma_{ij}(x, y).$$

Because of $f(\mathbf{x}) = g(\mathbf{x}, \overline{\mathbf{x}})$, we have for any $\mathbf{x} \in \mathcal{L}^n$:

$$h(\mathbf{x}) = \sum_{i \leq j} \eta_{ij}(x_i, x_j) = 0 \tag{23}$$

and therefore if we define θ as

$$\theta_{ij}(x, y) = \bar{\theta}_{ij}(x, y) - \eta_{ij}(x, y), \quad \theta'_{ij}(x, y) = \bar{\theta}'_{ij}(x, y),$$

it is a parameter set of g because of (9) and (23), and it satisfies (10). □

Proof (Theorem 2). By (12), we have

$$\alpha_{ij}(x_1, x_2, y_1, y_2) = \beta^{\theta}_{ij}(x_1, x_2, y_1, y_2) + \beta'^{\theta}_{ij}(x_1, x_2, y_1, y_2).$$

Thus the conditions (13) and (14) are equivalent to

$$\begin{aligned}\forall i \leq j, \forall x_1, x_2, y_1, y_2 \in \mathcal{L}, x_1 \prec x_2, y_1 \prec y_2 :& \\ \alpha^{-}_{ij}(x_1, x_2, y_1, y_2) \geq \beta^{\theta}_{ij}(x_1, x_2, y_1, y_2).&\end{aligned} \tag{24}$$

Moreover, for any $x_1 < x_2 < x_3, y_1 < y_2 < y_3$, we have

$$\begin{aligned}\beta^{\theta}_{ij}(x_1, x_3, y_1, y_3) = \beta^{\theta}_{ij}(x_1, x_2, y_1, y_2) + \beta^{\theta}_{ij}(x_1, x_2, y_2, y_3) +& \\ \beta^{\theta}_{ij}(x_2, x_3, y_1, y_2) + \beta^{\theta}_{ij}(x_2, x_3, y_2, y_3), \text{ and}&\end{aligned}$$

$$\begin{aligned}\alpha^{-}_{ij}(x_1, x_3, y_1, y_3) \geq \alpha^{-}_{ij}(x_1, x_2, y_1, y_2) + \alpha^{-}_{ij}(x_1, x_2, y_2, y_3) +& \\ \alpha^{-}_{ij}(x_2, x_3, y_1, y_2) + \alpha^{-}_{ij}(x_2, x_3, y_2, y_3).&\end{aligned}$$

Thus, the constraint $\alpha^{-}_{ij} \geq \beta^{\theta}_{ij}$ for the block (x_1, x_3, y_1, y_3) follows from that for the four blocks (x_1, x_2, y_1, y_2), (x_1, x_2, y_2, y_3), (x_2, x_3, y_1, y_2), and (x_2, x_3, y_2, y_3). In other words, the constraints for smaller blocks are more strict. Therefore, (24) is equivalent to the conditions for the smallest blocks:

$$\forall i \leq j : \forall x, y \in \{0, \ldots, l-2\} : \alpha^{-}_{ij}(x, x+1, y, y+1) \geq \beta^{\theta}_{ij}(x, x+1, y, y+1). \tag{25}$$

Thus, from (15), θ^* is a feasible solution to (11). By (25), the constraint

$$\begin{aligned}\forall i < j : \forall x_1, x_2, y_1, y_2 \in \mathcal{L}, x_1 \prec x_2, y_1 \prec y_2 :& \\ \sum_{x_1 \leq x < x_2} \sum_{y_1 \leq y < y_2} \alpha^{-}(x, x+1, y, y+1) \geq \beta^{\theta}_{ij}(x_1, x_2, y_1, y_2)&\end{aligned} \tag{26}$$

is also necessary. Because Lemma 2 allows us to assume $x_i < \overline{y}_i$ and $x_j < \overline{y}_j$,

$$g(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \sum_{i \leq j} \left(\beta^{\theta}_{ij}(x_i, \overline{y}_i, x_j, \overline{y}_j) + \gamma_{ij}(x_i, \overline{y}_j) + \gamma_{ij}(\overline{y}_i, x_j) \right)$$

is maximum if equality is attained on all constraints of (26). Thus, θ^* is optimal. □

References

1. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? IEEE Transactions Pattern Analysis and Machine Intelligence **24** (2004) 657–673 1
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions Pattern Analysis and Machine Intelligence **23** (2001) 1222–1239 1
3. Ishikawa, H.: Exact optimization for Markov random fields with convex prio rs. IEEE Transactions Pattern Analysis and Machine Intelligence **25** (2003) 1333–1336 1
4. Schlesinger, D., Flach, B.: Transforming an arbitrary minsum problem into a binary one. Technical report, TU Dresden (2006) 1, 8

5. Schlesinger, D.: Exact solution of permuted submodular minsum problems. In: Energy Minimization Methods in Computer Vision and Pattern Recognition. (2007) 28–38 1
6. Ishikawa, H.: Transformation of general binary MRF minimization to the first order case. IEEE Transactions Pattern Analysis and Machine Intelligence **33** (2011) 1234–1249 1
7. Gallagher, A.C., Batra, D., Parikh, D.: Inference for order reduction in Markov random fields. In: IEEE Conference on Computer Vision and Pattern Recognition. (2011) 1
8. Fix, A., Grubner, A., Boros, E., Zabih, R.: A graph cut algorithm for higher-order Markov random fields. In: IEEE International Conference on Computer Vision, Barcelona (2011) 1
9. Kahl, F., Strandmark, P.: Generalized roof duality for pseudo-boolean optimization. In: IEEE International Conference on Computer Vision. (2011) 1, 4
10. Kahl, F., Strandmark, P.: Generalized roof duality. Discrete Applied Mathematics (2012) 1, 4
11. Hammer, P.L., Hansen, P., Simeone, B.: Roof duality, complementation and persistency in quadratic 0-1 optimization. Math. Programming **28** (1984) 121–155 1
12. Boros, E., Hammer, P.L., Sun, R., Tavares, G.: A max-flow approach to improved lower bounds for quadratic unconstrained binary optimization (QUBO). Discrete Optimization **5** (2008) 501–529 1
13. Boros, E., Hammer, P.L., Tavares, G.: Preprocessing of unconstrained quadratic binary optimization. RUTCOR Research Report RRR 10-2006, Rutgers University (2006) 1
14. Rother, C., Kolmogorov, V., Lempitsky, V., Szummer, M.: Optimizing binary MRFs via extended roof duality. In: IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, Minnesota (2007) 1
15. Kolmogorov, V.: Generalized roof duality and bisubmodular functions. Discrete Applied Mathematics **160** (2012) 416 – 426 1
16. Kohli, P., Shekhovtsov, A., Rother, C., Kolmogorov, V., Torr, P.: On partial optimality in multi-label MRFs. In: Proceedings of the 25th international conference on Machine learning. (2008) 480–487 2, 8
17. Flach, B., Schlesinger, D.: Best labeling search for a class of higher order gibbs models. Pattern Recognition and Image Analysis **14** (2004) 249–254 4, 11