

Exhaustive Family of Energies Minimizable Exactly by a Graph Cut

Guillaume Charpiat
Pulsar team, INRIA
Sophia-Antipolis, France

Guillaume.Charpiat@inria.fr

Abstract

Graph cuts are widely used in many fields of computer vision in order to minimize in small polynomial time complexity certain classes of energies. These specific classes depend on the way chosen to build the graphs representing the problems to solve. We study here all possible ways of building graphs and the associated energies minimized, leading to the exhaustive family of energies minimizable exactly by a graph cut. To do this, we consider the issue of coding pixel labels as states of the graph, i.e. the choice of state interpretations. The family obtained comprises many new classes, in particular energies that do not satisfy the submodularity condition, including energies that are even not permuted-submodular.

A generating subfamily is studied in details, in particular we propose a canonical form to represent Markov random fields, which proves useful to recognize energies in this subfamily in linear complexity almost surely, and then to build the associated graph in quasilinear time. A few experiments are performed, to illustrate the new possibilities offered.

Introduction

Graph cut has been a very popular optimization tool in computer vision for a decade, because of their suitability to many problems defined on images, and because of their ability to find, in small polynomial time, global optima of energies which were mistakenly thought to be NP-hard to minimize before. The graph cut method consists in rewriting the energy minimization as a search for a minimal cut in a graph. The new problem obtained is known in graph theory as min-cut or maximum flow, and there exist efficient algorithms to solve it [5, 1], provided all weights in the graph are non-negative. The main question that this article addresses is to draw the list of all energies that can be minimized in such a way.

Other optimization techniques based on graphs exist, such as minimum ratio weight cycle, loopy belief propagation or spectral clustering. There might even exist other

techniques to estimate minimal cuts on different kinds of graphs [4]; this would however be out of the scope of this article, which focuses on graphs considered by usual minimal cut techniques, i.e. with positive weights.

Graph cuts were originally introduced in computer vision by [9] in 1989, and gained fame one decade later [18, 2]. Many new applications, performance improvements, wider classes of possible energies, or related optimization techniques then appeared [14, 23, 15, 21, 6, 8, 3]. In particular, the ability to solve certain classes of binary classification problems exactly with graph cuts was extended to the approximate resolution of multi-labeling problems with α -expansions or (α, β) -swaps [13], and even to exact resolution in the case of convex interaction terms between variables [11]. Since the class of energies solvable with graph cuts is regularly extended, an arising question is to know whether it would be possible to draw an exhaustive, definitive list which could not be broadened anymore. We will here focus on energies minimizable exactly, since including approximate techniques would involve their comparison and the quantization of their respective errors.

Most works on exact or approximate minimization with graph cuts require that pairwise interaction terms (restricted to any pair of labels) are submodular. However this submodularity condition is specific to particular ways of building graphs where each label is represented by one node, and we will see that, in the general case, the constraints are different. Amongst the energies that we are able to minimize, most are not submodular or not even permuted-submodular, and this is a major difference with classical graph construction and related studies [19, 22, 16, 20] about expressiveness or recognition of permuted problems, which once again focus on submodular energies or binary ones [7]. The only exceptions to the submodularity rule in the literature are based on *roof duality* [17], thanks to a different correspondence between labels and states of the nodes of the graph, but at the cost of giving only partial solutions to the problem, in the sense that they do not guarantee to assign a label to each unknown variable. We will be interested in all possible ways of coding label choices into graph partitions,

i.e. binary states of nodes. Once such a list is available, an issue arises, the one of recognizing whether an energy belongs to the list and of building the associated graph.

The article is organized as follows : In a first part, we discuss all possibilities to build positive-weighted graphs to represent energies to be minimized, and infer the exhaustive list \mathcal{F} of energies minimizable exactly by graph cuts. We then focus on a generating subfamily \mathcal{F}_0 . In the second part, we propose a canonical form to represent Markov Random Fields (MRF), in order to study \mathcal{F}_0 and express it in a more practical way. We will see that \mathcal{F}_0 includes new kinds of energies absent from the literature. In the third part we show how to recognize quickly energies of \mathcal{F}_0 (even after pixel-dependent permutations of labels). Finally, we perform a few experiments to illustrate the previous sections.

1. Graphs and interpretations

This section aims at drawing the list of all energies that can be rewritten as the cost of cuts in a graph with positive weights. Let $E(x_1, x_2, \dots, x_p)$ be a function of p variables x_i , to be minimized globally. Let G be a graph, consisting of q nodes N_j and of directed edges between pairs of nodes, with non-negative weights $w^{j \rightarrow k}$. Let us try to build E and G together so that cuts in the graph G can be interpreted as choices of values for all variables x_i .

If we restrict the study to the case of graphs with a finite number of nodes, then the number of possible cuts of a graph is finite, so the number of possible states of all variables $(x_i)_{1 \leq i \leq p}$ has to be finite. Hence each variable x_i has to be constrained to a finite set of possible values $X_i = \{L_i^1, L_i^2, L_i^3, \dots\}$ which we will name *labels*. The number of labels $\#X_i$ may depend on i .

1.1. Graph states

In order to use s - t min cut techniques such as max-flow [5], we introduce in the graph two special nodes, the sink and the source of the flow. It is possible to consider several sources or/and several sinks, but such a problem can be easily rewritten as a classical graph with only one source and one sink. The minimal cut obtained can be seen as a partition of the nodes of the graph : those which are connected to the source, and those which are connected to the sink. Thus, considering a cut is equivalent to picking a state $s(N_i)$ for each node of the graph amongst the two possibilities mentioned. Nodes can then be seen as binary variables, and graph cuts as states of all these binary variables together. Finding a minimal cut is then equivalent to finding a state s of the graph that solves:

$$s : \{N_i\}_{1 \leq i \leq q} \rightarrow \{\text{source, sink}\} \quad \min \sum_{i,j} w^{i \rightarrow j} \delta_{s(N_i) = \text{source} \ \& \ s(N_j) = \text{sink}}$$

where $\delta_P = 1$ if P is true, and 0 otherwise, thus summing over edges cut. Consequently, algorithms finding minimal

cuts [5, 1] can be used to find global minima of any energies of binary variables of the form

$$C(\mathbf{b}) = \sum_{i,j} w^{i \rightarrow j} \bar{b}^i b^j$$

where $\mathbf{b} \in \{0, 1\}^q$ and where $\bar{b}^i = 1 - b^i$, provided all coefficients $w^{i \rightarrow j}$ are non-negative. For any pair of binary variables $\{b^i, b^j\}$, their interaction cost matrix has consequently to be antidiagonal and positive, which leads to the well-known *submodularity* condition with usual graph constructions. Note however that this constraint applies to interaction costs between node states, and not to interaction costs between labels, that we have not defined yet.

1.2. Interpreting graph states as labels

We have to define a way to interpret graph states so that we can infer from any possible cut of the graph G , *i.e.* from any possible state of the binary variables b^j related to the nodes, the corresponding choice of labels for the variables $(x_i)_{1 \leq i \leq p}$ of the energy E . Let us consider one variable x_i : its label has to be a function of the states of the nodes of the graph. This function might depend on a few nodes only; let us denote by S_i the set of nodes upon which x_i depends, it is a subpart of G . This set S_i of binary variables can be ordered arbitrarily : $(b_i^1, b_i^2, \dots, b_i^{q_i})$, leading to a numbering $\llbracket 1, 2^{q_i} \rrbracket$ of all possible states of the subgraph S_i as integers expressed in base 2. We will name *interpretation* any function from this numbering to the set X_i of labels available for the energy variable x_i , *i.e.* any function

$$I_i : \llbracket 1, 2^{q_i} \rrbracket \longrightarrow \llbracket 1, \#X_i \rrbracket.$$

1.2.1 The simple case

Let us first assume that each node of the graph G belongs to exactly one subgraph S_i , *i.e.* that the subgraphs S_i form

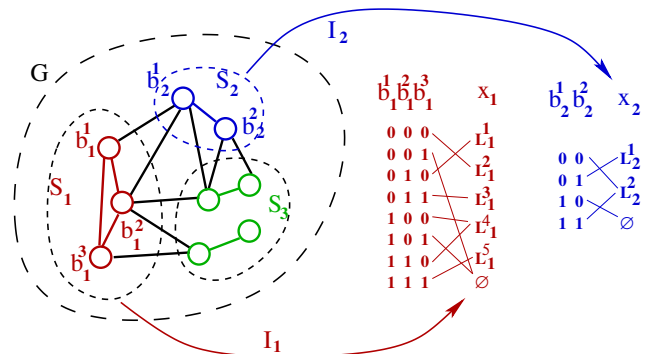


Figure 1. The states of subparts S_i of the graph G can be represented by a set of binary variables b_i^j , which are then interpreted together as labels L_i^k of the energy variable x_i via the *interpretation* function I_i . Some states of the subgraph S_i may never be reached, in that case they do not need to correspond to a label (\emptyset).

a partition of G , as in Figure 1. Then, let us be given a choice of interpretations I_i for all variables x_i . For the sake of readability, we will suppose in a first step that each interpretation I_i is a one-to-one correspondence between subgraph states $(b_i^1, b_i^2, \dots, b_i^{q_i})$ and possible labels of variable x_i . Then the binary variable b_i^k can be seen as a function of x_i : it is the k^{th} digit of $I_i^{-1}(x_i)$. Thus, the energy $C(\mathbf{b})$ rewrites as a function $E(x)$ of the variables x_i , by regrouping interactions between binary variables by subgraphs:

$$E(x) = \sum_i D_i(x_i) + \sum_{i,j} V_{i,j}(x_i, x_j) \quad (1)$$

where

$$D_i(x_i) = \sum_{k=1}^{q_i} \left(w_i^{\text{source} \rightarrow k} b_i^k + w_i^{k \rightarrow \text{sink}} \overline{b_i^k} \right) + \sum_{k=1}^{q_i} \sum_{\substack{k'=1 \\ k' \neq k}}^{q_i} w_i^{k \rightarrow k'} \overline{b_i^k} b_i^{k'}$$

and

$$V_{i,j}(x_i, x_j) = \sum_{k=1}^{q_i} \sum_{l=1}^{q_j} w_{i,j}^{k \rightarrow l} \overline{b_i^k} b_j^l + w_{i,j}^{k \leftarrow l} b_i^k \overline{b_j^l}$$

where $w_i^{k \rightarrow k'}$ stands for the weight of the edge from the k -th to the k' -th node in subgraph S_i , and similarly $w_{i,j}^{k \rightarrow l}$ for edges between different subgraphs. The source and the sink were distinguished from other nodes.

Definition 1 : Subfamily \mathcal{F}_0 . We will denote by \mathcal{F}_0 the family of all energies that can be written in such a way, for any finite number of variables x_i , any finite sets of labels X_i , any one-to-one interpretations I_i and any non-negative coefficients w . Thus \mathcal{F}_0 is the set of all energies minimizable by a cut in a graph, up to the two previous assumptions, that (S_i) is a partition of G , and that $\forall i, I_i$ is one-to-one.

1.2.2 Sharing nodes; nodes unused in labeling

The first hypothesis made was that the subgraphs S_i formed a partition of the graph G . If instead several subgraphs share a same node, then this node may be divided into as many instances as subgraphs, with infinite links (in both directions) between the instances, so that each subgraph can have its own copy of the node without sharing it, while all copies share a common state. The problem of overlapping subgraphs is thus reduced to the previous case. On the other hand, there may be nodes in G which do not belong to any subgraph S_i , i.e. their states are not taken into account in any interpretation x_i . The set of all such nodes forms another subgraph S_{supp} , which can be seen as a supplementary variable x_{supp} . Then we have a partition of the graph and can apply the previous approach. We will obtain an energy of the same form $E_1(x, x_{\text{supp}})$; since we are interested in minimal cuts, one can consider that the quantity minimized is the infimum over x_{supp} of this energy : $E_2(x) = \min_{x_{\text{supp}}} E_1(x, x_{\text{supp}})$.

1.2.3 Impossible states; multiply-represented labels

Some states of a subgraph S_i might be non-reachable because they correspond to infinite costs; this happens when infinite weights are set between nodes of the same subgraph S_i , or between one of its nodes and the source or the sink. The interpretation of such states has no consequence since they are never reached; one can choose to map them to any possible label of x_i , to map them to an additional explicit label \emptyset (which will never be chosen), or just to ignore this state when defining interpretations. On the opposite, several different states of a subgraph S_i may be associated with a same label L_i^k of the variable x_i . Then the cost of choosing this label will be the minimum of the costs of the different possible interpretations. Let us consider the same graph but with a one-to-one interpretation, with supplementary labels so that each state of each subgraph corresponds to exactly one label, and denote by $X(L_i^k)$ the set of new labels corresponding to the former label L_i^k . Then we can define : $E'(x) = \min_{y \text{ s.t. } y_i \in X(x_i)} E(y)$, where $E \in \mathcal{F}_0$. This energy E' satisfies that its global minimum is reached for a minimal cut on G .

1.2.4 Complete family

Now that we have removed the only two assumptions made earlier, we obtain :

Proposition 1 : Exhaustive family. The exhaustive family \mathcal{F} of all energies minimizable by a graph cut is the set of functions which either belong to the former family \mathcal{F}_0 , either can be written as an infimum of an energy in \mathcal{F}_0 with respect to some of its variables, globally (i.e. over all possible labels of these variables : $\min_{x_{\text{supp}}}$) or partially (i.e. over a subset of the possible labels of these variables : $\min_{x_i \in X_i^r \subset X_i}$).

1.3. Known methods as particular cases

To see how standard graph constructions are included in \mathcal{F} , we summarize in Fig. 2 the main different approaches to global minimization by graph cuts in the literature. We indicate the number of labels, of nodes per subgraph, as well as the interpretation function chosen and other particularities.

2. Rewriting in a canonical form

We now aim at expressing energies in the families \mathcal{F} or \mathcal{F}_0 in a more suitable way than equation (1) and its constraints on admissible functions D_i and $V_{i,j}$. For this, we will first need to study the space of energies related to Markov random fields.

2.1. Markov random fields

Let \mathcal{F}_M be the set of all energies E of the kind :

$$E(x) = \sum_i D_i(x_i) + \sum_{i,j} V_{i,j}(x_i, x_j) \quad (2)$$

Method	$\#X_i$	$\#S_i$	I_i	Notes
Binary graph cut	$2 \forall i$	1	$Id : \begin{matrix} 0 \mapsto 0 \\ 1 \mapsto 1 \end{matrix}$	Submodularity required
Ishikawa [11]	$L \forall i$	L	bijection node-label (fixed numbering)	Infinite weights between nodes to maintain order; Convex interaction term
Roof duality, a.k.a. QPBO [17]	$3 \forall i:$ $\{0, 1, '?'\}$	2	$\begin{matrix} (0, 1) \mapsto 0 \\ (1, 0) \mapsto 1 \\ (0, 0), (1, 1) \mapsto '?' \end{matrix}$	Partial answer : '?' = unknown label; Allows supermodularity on a binary problem
Generating family \mathcal{F}_0 (this work)	2^{n_i}	n_i	any bijection	Number of labels = variable-dependent; Includes non permuted-submodular energies; Constraints checked in linear time

Figure 2. Most common approaches to global minimization with graph cuts in the literature, and how they can be understood in this framework. Number of labels per variable, number of nodes per subgraphs are indicated, as well as the interpretation function.

where each variable x_i belongs to a finite set of possible labels X_i , and where D_i and $V_{i,j}$ are any real-valued functions, of one or two of these variables respectively. Such energies are related to *Markov random fields* (MRF). Note that $\mathcal{F}_0 \subset \mathcal{F}_M$.

Given an ordering of the possible labelings X_i of each variable x_i , the functions $D_i(\cdot)$ and $V_{i,j}(\cdot, \cdot)$ can be respectively represented as vectors and matrices with real coefficients, since they are real-valued functions defined on a finite ordered set or on a product of two finite ordered sets. The k^{th} component of the vector D_i is the local cost $D_i(L_i^k)$ of choosing the k^{th} possible label L_i^k for variable x_i , and similarly the $(k, l)^{\text{th}}$ coefficient of matrix $V_{i,j}$ is the interaction cost $V_{i,j}(L_i^k, L_j^l)$ between the k^{th} possible label for variable x_i and the l^{th} possible label for x_j .

2.2. Canonical form

In general, there are many different ways to express a labeling cost in the form of equation (2), so that two energies may have different vectors D_i and matrices $V_{i,j}$ while leading to the same costs for all possible labelings. For example, the following two Markov chains C_1 and C_2 are equivalent :

$$E = D_1(x_1) + V_{12}(x_1, x_2) + D_2(x_2) + V_{23}(x_2, x_3) + D_3(x_3)$$

$$C_1 \begin{pmatrix} 2 \\ 3 \end{pmatrix} \begin{pmatrix} 4 & 7 \\ 5 & 0 \end{pmatrix} \begin{pmatrix} -4 \\ 1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$C_2 \begin{pmatrix} 6 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 8 & 0 \end{pmatrix} \begin{pmatrix} -7 \\ 3 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

We will now rewrite any energy in \mathcal{F}_M so that the sums of elements of each row, each column of each matrix $V_{i,j}$ are 0. To do this, let \mathcal{M}_S be the vector space of matrices of the form $M = (r_l + c_k)_{k,l}$. Such matrices are a sum of a matrix with identical rows r and of a matrix with identical columns c . We project $V_{i,j}$ orthogonally on \mathcal{M}_S to obtain the row $r^{i,j}$ and the column $c^{i,j}$, and we note $V_{i,j}^0$ the remaining

orthogonal part, i.e. $V_{i,j}^0 = V_{i,j} - (r_l^{i,j} + c_k^{i,j})_{k,l}$. Let

$$D_i^{\text{new}} = D_i + \sum_j c^{i,j} + \sum_j r^{j,i}. \quad (3)$$

The system obtained (D^{new}, V^0) expresses the same costs as the original one (D, V) . We also subtract the mean values of each vector D_i^{new} ; the system (D^0, V^0) thus obtained expresses the same energy functional as (D, V) , up to a constant.

Proposition 2 : Canonical form. *Any MRF energy (i.e., in \mathcal{F}_M) $E = \sum_i D_i + \sum_{i,j} V_{i,j}$ can be rewritten, up to a constant, as $\sum_i D_i^0 + \sum_{i,j} V_{i,j}^0$ where the mean value of each vector D_i^0 is 0, as well as the mean value of each line, each column of each matrix $V_{i,j}^0$. This writing is unique and will be referred as the canonical form of E .*

Proof: Existence by construction. Uniqueness : Let A and B be two writings of the same energy in canonical forms. They are equal (up to a constant shift) for all labelings; let us express them as vectors and matrices for a common ordering of labels, and write their difference δE with vectors $D_i = D_i^A - D_i^B$ and matrices $V_{i,j} = V_{i,j}^A - V_{i,j}^B$. We want to show that these vectors and matrices are 0. By hypothesis, the function $\delta E(x)$ is a constant and does not depend on x . Let us consider one particular variable x_i and fix other variables x_j . The sum of the terms involving x_i is $D_i(x_i) + \sum_j V_{i,j}(x_i, x_j)$ and does not depend on x_i (since δE is constant). If we sum this quantity over possible values $x_i \in X_i$, we obtain $\#X_i$ times the same value, but also 0 (mean values of vectors or columns are 0) so that $D_i(x_i) + \sum_j V_{i,j}(x_i, x_j) = 0 \forall i, x_i, x_j$. If we now average this equality over all values $x_j \in X_j$ for all $j \neq i$, we find $D_i(x_i) = 0$ (since row means are 0), so that $\sum_j V_{i,j}(x_i, x_j) = 0 \forall i, x_i, x_j$. Then $\forall i, j, x_i$, the only term of this constant sum that involves x_j is $V_{i,j}(x_i, x_j)$ and cannot depend on x_j . Since moreover the (row) mean of $V_{i,j}(x_i, x_j)$ over x_j is 0, we have $V_{i,j}(x_i, x_j) = 0$. Consequently, $D^A = D^B$ and $V^A = V^B$, hence uniqueness. \square

2.3. The energies of the family, in canonical form

Energies of the family \mathcal{F}_0 are written in (1) as linear combinations of quantities b_i^k , $\overline{b_i^k} b_i^{k'}$ and $\overline{b_i^k} b_j^l$, which can be seen as vectors or matrices as functions of the subgraph states representing x_i and x_j . Their canonization, detailed in [24], leads respectively to vectors $A_i^k, A_i^{k,k'}$ and matrices $W_{i,j}^{k,l}$ defined by $(A_i^k)_s = (-1)^{1+b^k(s)}$, $(A_i^{k,k'})_s = (-1)^{1+b^k(s)+b^{k'}(s)}$, and $W_{i,j}^{k,l} = -A_i^k \otimes A_j^l$, with $b^k(s)$ standing for the k^{th} bit of the binary writing of the state s . For a given i , the vectors A_i^k and $A_i^{k,k'}$ form an orthogonal family of $q_i(q_i + 1)/2$ elements. For given i and j , the matrices $W_{i,j}^{k,l}$ form an orthogonal family too, of $q_i q_j$ elements.

We now rewrite \mathcal{F}_0 in a more practical way. Let E be an energy in \mathcal{F}_0 as in equation (1) with satisfied constraints, with a given ordering of the nodes in each subgraph. The canonical form of E is then :

$$\left\{ \begin{array}{l} D_i^0 = \sum_{k=1}^{q_i} \alpha_i^k A_i^k + \sum_{k=1}^{q_i} \sum_{\substack{k'=1 \\ k' \neq k}}^{q_i} \alpha_i^{k,k'} A_i^{k,k'} \\ V_{i,j}^0 = \sum_{k=1}^{q_i} \sum_{l=1}^{q_j} \alpha_{i,j}^{k,l} W_{i,j}^{k,l} \end{array} \right. \quad (4)$$

where

$$\left\{ \begin{array}{l} \alpha_i^k = \frac{1}{2}(w_i^{\text{source} \rightarrow k} - w_i^{k \rightarrow \text{sink}}) + \frac{1}{4} \sum_{k'} (w_i^{k \leftarrow k'} - w_i^{k \rightarrow k'}) \\ \quad + \frac{1}{4} \sum_j (w_{i,j}^{k \leftarrow k'} - w_{i,j}^{k \rightarrow k'}) \\ \alpha_i^{k,k'} = \frac{1}{4}(w_i^{k \rightarrow k'} + w_i^{k \leftarrow k'}) \\ \alpha_{i,j}^{k,l} = \frac{1}{4}(w_{i,j}^{k \rightarrow l} + w_{i,j}^{k \leftarrow l}) \end{array} \right.$$

can be seen as parameters of the energy E , *i.e.* as degrees of freedom in the design of energies within \mathcal{F}_0 .

Note: no degree of freedom is lost by assuming $w_i^{k \rightarrow k'} = w_i^{k \leftarrow k'}$ and $w_{i,j}^{k \rightarrow l} = w_{i,j}^{k \leftarrow l}$, thanks to $w_i^{\text{source} \rightarrow k}$ and $w_i^{k \rightarrow \text{sink}}$ which make the span of α_i^k to be fully \mathbb{R} . Thus it is sufficient to consider only undirected graphs G .

Proposition 3 : Characterization of \mathcal{F}_0 . *The family \mathcal{F}_0 is the set of MRF energies such that there exist an ordering of the labels, and parameters $\alpha_i^k \in \mathbb{R}$ and $\alpha_i^{k,k'}, \alpha_{i,j}^{k,l} \in \mathbb{R}^+$, for which the canonical expression $(D_i^0, V_{i,j}^0)$ of the energy satisfies equations (4).*

Let us detail the vectors $A^k, A^{k,k'}$ and the matrices $W^{k,l}$ which generate the span of energies minimizable by a graph cut. If a subgraph contains only $\#S_i = 1$ node, then $A^1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ and since $\alpha^1 \in \mathbb{R}$, all canonized potentials D_i^0 of a

binary variable are representable. In the case of $\#S_i = 2$ nodes :

$$A_i^1 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix}, \quad A_i^2 = \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \quad A_i^{1,2} = \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

and since the dimension of the space of canonized potentials is 3, with $\alpha^1, \alpha^2 \in \mathbb{R}$ and $\alpha^{1,2} \in \mathbb{R}^+$, half of them are representable for any fixed interpretation. For $\#S_i = 3$:

$$\begin{array}{cccccc} \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, & \begin{pmatrix} -1 \\ -1 \\ 1 \\ -1 \\ -1 \\ 1 \\ 1 \end{pmatrix}, & \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ 1 \end{pmatrix}, & \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, & \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, & \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} \\ A_i^1 & A_i^2 & A_i^3 & A_i^{1,2} & A_i^{1,3} & A_i^{2,3} \end{array}$$

6 degrees of freedom are available (amongst 7). Concerning interaction matrices : for a pair of subgraphs containing one node each, there is only one interaction matrix generator :

$$W_{1,2}^{1,1} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The space of all canonized interaction matrices M (for any MRF, not necessarily in \mathcal{F}_0) has only one degree of freedom, so that any interaction matrix M writes necessarily as $\alpha W_{1,2}^{1,1}$ for an $\alpha \in \mathbb{R}$. If $\alpha \geq 0$, M is said to be submodular, if $\alpha \leq 0$, M is supermodular. Since $\alpha^{1,1}$ has to be positive, any energy of binary variables which is in \mathcal{F}_0 has to satisfy that there exists a label ordering for which $V_{i,j}^0$ is submodular for all (i, j) simultaneously. There are however other interpretations for which given $V_{i,j}^0$ are supermodular.

The matrices available in \mathcal{F}_0 for the interaction between a subgraph of 1 node and one of 2 nodes are :

$$\begin{array}{cc} \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}, & \begin{pmatrix} -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \\ W_{1,2}^{1,1} & W_{1,2}^{1,2} \end{array}$$

For interactions between two subgraphs of 2 nodes each :

$$\begin{array}{cccc} \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}, & \begin{pmatrix} -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}, & \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}, & \begin{pmatrix} -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \\ W_{1,2}^{1,1} & W_{1,2}^{1,2} & W_{1,2}^{2,1} & W_{1,2}^{2,2} \end{array}$$

Note that in the three last matrices, there exist 2×2 submatrices which are supermodular. Moreover, many positive linear combinations of these matrices are **not permuted-submodular**, *i.e.* there exists no interpretation (= permutation) that makes all 2×2 submatrices become submodular. Consider *e.g.* $\alpha^{1,1} = \alpha^{2,2} = 1/2$, $\alpha^{2,1} = \alpha^{1,2} = 0$:

$$\begin{pmatrix} -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}. \text{ This proves that the span of energies}$$

minimizable globally with graph cuts is much wider than presented in the literature since there was no graph construction able to deal with non permuted-submodularity so far, or even just non-submodular multilabel cases.

If we replace '1' with '+', and '-1' with an empty space to better visualize matrices, we get, for interactions between two subgraphs of 3 nodes each, the 9 following matrices :

$$\begin{pmatrix} & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{pmatrix}, \begin{pmatrix} & + & + & & + & + \\ & + & + & & + & + \\ & + & + & & + & + \\ & + & + & & + & + \\ + & + & + & + & + & + \\ + & + & + & + & + & + \\ + & + & + & + & + & + \end{pmatrix}, \begin{pmatrix} & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{pmatrix},$$

$$\begin{pmatrix} & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{pmatrix}, \begin{pmatrix} & + & + & & + & + \\ & + & + & & + & + \\ & + & + & & + & + \\ & + & + & & + & + \\ + & + & + & + & + & + \\ + & + & + & + & + & + \\ + & + & + & + & + & + \end{pmatrix}, \begin{pmatrix} & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{pmatrix},$$

$$\begin{pmatrix} & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{pmatrix}, \begin{pmatrix} & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{pmatrix}, \begin{pmatrix} & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{pmatrix}.$$

We note that the matrices $W^{k,l}$ form an orthogonal family of centrally-symmetric Haar-like wavelets. They encode all possible regular patterns '+' with power-of-2 frequencies and form a regular grid in Fourier space with log-scale axes.

3. Recognizing efficiently energies in \mathcal{F}_0

Changing an ordering of labels I_i for another one can be seen as a coherent permutation of the rows and/or columns of the vectors D_i and matrices $V_{i,j}$. Orderings I_i depend on i , i.e. the interpretations of two different variables generally differs. For an MRF with p variables with L possible labels each, there are L^p possible orderings.

Proposition 4 : Checking \mathcal{F}_0 membership and finding orderings. *The complexity of testing whether an MRF energy belongs to \mathcal{F}_0 , and, in that case, of finding an ordering of labels $(I_i)_i$ so that it can be expressed as in Proposition 3, is linear as a function of the data size almost surely.*

Proof: By construction, as follows. By "data size" we mean the number of coefficients in vectors and matrices describing the energy E , and "almost surely" means that the measure of the set of energies for which this is not true is 0 when coefficients are taken randomly in \mathbb{R} or \mathbb{R}^+ .

Let us suppose that E belongs to \mathcal{F}_0 and writes as in Prop. 3 for the interpretation $(I_i)_i$, which we are searching for. Let us focus on a particular interaction matrix $N = V_{i,j}^0$ between two variables, and let M be its writing with the good interpretations (I_i, I_j) . N is a permuted version of M , and we aim at guessing I_i and I_j from just N . The coefficients of M are $M_{s_1, s_2} = \sum_{k,l} \alpha^{k,l} (-1)^{1+b^k(s_1)+b^l(s_2)}$. The

smallest one is $-\sum_{k,l} \alpha^{k,l}$, reached for $s_1 = s_2 = 0000\dots$ and for $s_1 = s_2 = 1111\dots$ in binary writing (the first and last elements of the matrix). They are reached nowhere else provided all $\alpha^{k,l} > 0$. Since this value does not vary with permutations, we can identify in N the rows/columns corresponding to the first and last ones of M . Let us now study the first row $R_s = M_{000\dots, s}$. We note that $R_s = -\sum_k (\sum_l \alpha^{k,l}) (-1)^{b^k(s)} =: -\sum_k (-1)^{b^k(s)} \beta_k$ with $\beta_k \geq 0$ (one per node). Then, if we know the (β_k) , we can infer the binary label s from just the coefficient value R_s by generating all combinations $\sum_k \pm \beta_k$ and check which one equals R_s . This works if all R_s are different (which is the case almost surely if α are taken randomly in \mathbb{R}^+ for any standard measure). Thus we can re-order the first row after finding it scrambled, and thus, with the same process on the first column, we have un-permuted the whole matrix N , and the problem is solved. Thus, we only need now to find the $\{\beta_k\}$ from just the set of scrambled values $\{R_s\}$. Note that the numbering of nodes in the subgraph S_i related to variable x_i is not relevant (it does not change the graph nor the energy), and that it acts on matrices $\{W^{k',l'}\}$ by swapping them, and similarly on $\{\beta_k\}$. We chose to order canonically the nodes by decreasing order of their β_k . Also, M is centrally symmetric, so all results are up to an order reversing permutation. We now present two algorithms with same complexities to estimate the $\{\beta_k\}$ from the $\{R_s = \sum_k \pm \beta_k\}$. The first one is intriguing but works only almost surely, while the second one always does.

The difference frequency algorithm consists in computing all pairwise differences $|R_s - R_{s'}|$ and in sorting the values obtained in an histogram. Then the $\#S_i$ most frequent values are precisely the $\beta_k (\times 2)$, provided the sums over subsets $\sum_{k \in K} \beta_k$ are all different. The complexity is the square of the number of labels, which is the size of the interaction matrix. \square

The $+\beta_k$ graph algorithm consists in constructing iteratively a graph whose nodes are values R_s and whose edges link two values whose difference is of the form $2\beta_k$. The iteration is done by exploring the remaining values R_s sorted in increasing order in a buffer, and the values β_k are guessed in increasing order. We start the graph with one node, $R_{min} = \min_s R_s$. At step t , the t smallest β_k are known, and the graph consists of all combinations $R_s = \sum_n \pm \beta_n$ with negative signs for all β_n that are not amongst the t smallest ones. At each step, we first remove the next smallest R_s in the buffer and add it to the graph. Its difference with R_{min} is precisely the next smallest $\beta_k (\times 2)$, because all R_s whose difference with R_{min} is a combination of smaller $\beta_i (\times 2)$ are already in the graph. Second, for each already existing node R_{old} in the graph (except the last one), we search the buffer for exactly one R_s which is the sum of R_{old} and of the last found $2\beta_k$, and place it in the graph. Then we iterate. Complexity : as previously. \square

With the process above, the matrix $V_{i,j}^0$ gives a.s. the unique possible interpretation of variables x_i and x_j up to the group of node renumbering and to order reversion. Checking that all interpretations from all matrices are compatible together (same orbit) as well as with potentials D_i^0 takes linear time.

Corollary (uniqueness): *The writing of an energy in \mathcal{F}_0 as in Proposition 3 is almost surely unique, i.e. there exists a.s. only one unique valid interpretation (= label orderings), up to node numbering and global reversion of all orderings.*

Then, using the orthogonal decomposition on the bases A or W , the complexity of finding all coefficients is $p\mathcal{N}L^2(\log_2 L)^2$ if \mathcal{N} is the average neighborhood size. It is to be compared with the number of parameters $\alpha : p\mathcal{N}(\log_2 L)^2$ and to the total data size required to describe E (coefficients of the MRF) : $p\mathcal{N}L^2$. The total complexity is thus linear times $(\log_2 L)^2$, and thus surprisingly low.

Proposition 5 : Building graph. *The construction of a graph related to an energy in \mathcal{F}_0 can be done in quasilinear time as a function of the data size.*

Note that given an energy in \mathcal{F}_0 , the required number of nodes in the graph to encode the state of one variable x_i with L possible labels is only $\log_2 L$. Thus the graphs are small, and so is the complexity of the graph cut, based on the numbers of nodes $p \log_2 L$ and of edges $p\mathcal{N}(\log_2 L)^2$.

4. Examples : enforcing edges with \mathcal{F}_0

Many computer vision problems involve Markov random fields, where the family \mathcal{F}_0 could be of practical interest. For instance in image segmentation, the detection of edges is usually more robust than the classification of pixels based on local appearance (e.g., local descriptors such as texture), which means that an important part of the information is contained in the interaction matrices $V_{i,j}^0$. However, under a submodularity assumption as with classical graph-cut techniques, it is not possible to impose variations of labels where edges are detected, if one does not know in advance which labels are on each side of the edge. This is an important limitation. On the contrary, the family \mathcal{F}_0 is much wider, and we are able to enforce edges in problems we know to minimize globally. Yet, not all possible individual potentials and interaction matrices are available (see [12] to be able to pick any potential). To illustrate the possibilities offered, we show examples of problems in \mathcal{F}_0 (not meant as real applications), involving 4 labels in Fig. 3, or 8 labels in Fig. 4. The latter requires to replace the data-defined potentials D_i^0 with admissible ones in \mathcal{F}_0 , by projection on the orthogonal family $A^k, A^{k,k'}$. Table 1 displays a few possible 8×8 interaction matrices.

Discussion

We have drawn and studied the exhaustive family \mathcal{F} of all energies whose global minimum can be computed by a $s-t$ min cut, e.g. with a max-flow algorithm, in small poly-

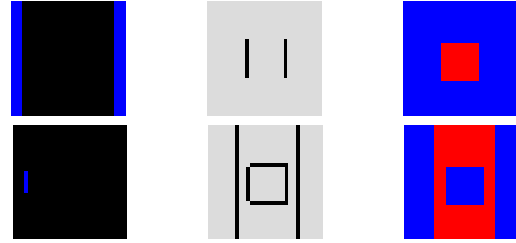


Figure 3. Examples of 2 non-submodular problems within the family \mathcal{F}_0 (one per row). We consider 4 possible labels : Green, Blue, Red, Purple. On the left, the potentials (the blue zones prefer to be blue, with $\alpha^1 = -\alpha^2 = 1/2$, and the black zones have no preference: $\alpha^i = 0$). In the middle, the interaction term between neighbors: in grey, coherency ($\alpha^{1,1} = \alpha^{2,2} = 1$, $\alpha^{1,2} = \alpha^{2,1} = 0$), and in black, change of labels ($\alpha^{1,1} = \alpha^{2,2} = 0$, $\alpha^{1,2} = \alpha^{2,1} = 10$). The corresponding potential vectors and interaction matrices are : $\begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, and $2 \begin{pmatrix} -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}$ and $20 \begin{pmatrix} -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}$. On the right, the result. If as here the first and last labels are not desirable, the problem is binary (Red vs. Blue) and the answer may be partial (when another label is chosen). This particular formulation boils down to QPBO indeed. Note however that, on the opposite to QPBO, we can apply the same trick with any number of nodes and labels, for instance with 8 labels as in Fig. 4.

mial time. By definition, this list includes all classes known to be minimizable exactly by graph cuts in the literature, but also new energies, in particular non-submodular ones. The family \mathcal{F} is generated by a simpler one \mathcal{F}_0 in the sense that elements of \mathcal{F} write as partial or global infimum of an energy in \mathcal{F}_0 (prop. 1) over some of its variables. Thanks to a canonical way of representing Markov random fields with vectors and matrices (prop. 2), we expressed the family \mathcal{F}_0 in a practical way as a subset of a vector space with known generating elements, up to permutations of rows and columns (prop. 3). We then showed how to optimize over all possible permutations in order to recognize any element in the generating family \mathcal{F}_0 with linear complexity a.s. (prop. 4), and how to rewrite it in a suitable way for graph cut minimization, in almost linear complexity. We also stressed the importance of enforcing edges during image segmentation, where \mathcal{F}_0 might offer new possibilities of practical interest.

The study of the complete family \mathcal{F} based on the generating subfamily \mathcal{F}_0 remains to be done, in the same spirit as classes of minimizable binary energies are generated from the set of binary submodular energies in the literature.

Since the MRF class of energies contains mainly NP-hard problems, most energies will not be solvable by a graph cut. Thus a promising track of research is the approximation of any MRFs [10] with energies in \mathcal{F} or \mathcal{F}_0 .

Our code is written in C++ and is available online at [24].

Acknowledgments

The author would like to thank Mickaël Péchaud for having convinced him that graph cut is a very general, powerful optimization tool (and not just a fashionable trend in vision).



Figure 4. An example of image digitalization in 8 grey levels, with edges enforced, in the same spirit as in Figure 3. We consider 8 classes of grey levels. The individual potentials rely on the intensity of pixels. The coherency criterion makes use of an edge detector and enforces label variations on edges. From left to right : most probable grey level (at each pixel independently) of the original potentials; of the potentials once projected on the set of admissible potentials (*i.e.* on \mathcal{F}_0); edges enforced; result. It is possible to let the user interact with the process and add edges manually to obtain new digitalization results. The interaction matrices used are (a) and (b) in Table 1.

$$\begin{pmatrix} -3 & -1 & -1 & 1 & -1 & 1 & 1 & 3 \\ -1 & -3 & 1 & -1 & 1 & -1 & 3 & 1 \\ -1 & 1 & -3 & -1 & 1 & 3 & -1 & 1 \\ 1 & -1 & -1 & -3 & 3 & 1 & 1 & -1 \\ -1 & 1 & 1 & 3 & -3 & -1 & -1 & 1 \\ 1 & -1 & 3 & 1 & -1 & -3 & 1 & -1 \\ 1 & 3 & -1 & 1 & -1 & 1 & -3 & -1 \\ 3 & 1 & 1 & -1 & 1 & -1 & -1 & -3 \end{pmatrix} \quad (a)$$

$$\begin{pmatrix} -3 & -1 & -1 & 1 & -1 & 1 & 1 & 3 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 3 & 1 & 1 & -1 & 1 & -1 & -1 & -3 \end{pmatrix} \quad (b)$$

$$\begin{pmatrix} -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 \end{pmatrix} \quad (c)$$

$$\begin{pmatrix} -2 & 0 & -1 & 1 & -1 & 1 & 0 & 2 \\ 0 & 2 & -1 & 1 & -1 & 1 & -2 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 \\ -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & -2 & 1 & -1 & 1 & -1 & 2 & 0 \\ 2 & 0 & 1 & -1 & 1 & -1 & 0 & -2 \end{pmatrix} \quad (d)$$

Table 1. A few 8×8 interaction matrices available in \mathcal{F}_0 . Matrix (a) expresses coherency (labels prefer not to change : very negative diagonal), while (b) prefers label changes (except for the first and the last labels, positive diagonal). Other kinds of label variations can be preferred, such as (c) or (d), *etc.*, depending on the label groups between which edges are wished to be valid. Matrix (a) is obtained with $\alpha^{1,1} = \alpha^{2,2} = \alpha^{3,3} = 1$ while other $\alpha^{k,l} = 0$. For matrix (b), on the contrary, $\alpha^{1,1} = \alpha^{2,2} = \alpha^{3,3} = 0$ while other $\alpha^{k,l} = 1/2$. For matrix (c) : $\alpha^{1,2} = \alpha^{2,1} = 1/2$ while others are 0. For matrix (d) : $\alpha^{1,2} = \alpha^{2,1} = \alpha^{1,3} = \alpha^{3,1} = 1/2$ while others are 0.

References

- [1] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *TPAMI*, (26(9)):1124–1137, 2004. 1, 2
- [2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In *ICCV*, 1999. 1
- [3] P. Carr and R. Hartley. Solving multilabel graph cut problems with multilabel swap. In *Digital Image Computing: Techniques and Applications*, pages 532–539, 2009. 1
- [4] A. Coja-Oghlan. Solving NP-hard semirandom graph problems in polynomial expected time. *J. Algorithms*, 2007. 1
- [5] G. B. Dantzig and D. R. Fulkerson. On the max-flow min-cut theorem of networks. *Ann. Math. Studies*, (38), 1956. 1, 2
- [6] J. Darbon. Global optimization for first order Markov random fields with submodular priors. In *Int. Workshop on Combinatorial Image Analysis*, 2008. 1
- [7] D. Freedman and P. Drineas. Energy minimization via graph cuts: Settling what is possible. In *CVPR*, 2005. 1
- [8] S. Gould, F. Amat, and D. Koller. Alphabet soup: A framework for approximate energy minimization. In *CVPR'09*. 1
- [9] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *J. R. Statist. Soc. B*, (51):271–279, 1989. 1
- [10] G. Hamarneh. Multi-label MRF optimization via a least squares s-t cut. In *ISVC*, 2009. 7
- [11] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *TPAMI*, 25, October 2003. 1, 4
- [12] V. Kolmogorov. A dichotomy theorem for conservative general-valued CSPs. Technical report, 2010. 7
- [13] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts ? In *ECCV (3)*, 2002. 1
- [14] N. Komodakis and G. Tziritas. A new framework for approximate labeling via graph cuts. In *ICCV*, 2005. 1
- [15] V. Lempitsky, C. Rother, and A. Blake. Logcut - efficient graph cut optimization for Markov random fields. In *ICCV*, Rio de Janeiro, Brazil, 2007. 1
- [16] S. Ramalingam, P. Kohli, K. Alahari, and P. H. S. Torr. Exact inference in multi-label CRFs with higher order cliques. In *CVPR*, 2008. 1
- [17] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *CVPR*, 2007. 1, 4
- [18] S. Roy and I. J. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *ICCV*, 1998. 1
- [19] D. Schlesinger. Exact solution of permuted submodular min-sum problems. In *EMMCVPR'07*, 2007. 1
- [20] D. Schlesinger and B. Flach. Transforming an arbitrary min-sum problem into a binary one. Technical report, 2006. 1
- [21] O. Veksler. Graph cut based optimization for MRFs with truncated convex priors. In *CVPR*, 2007. 1
- [22] S. Žitný, D. A. Cohen, and P. G. Jeavons. The expressive power of binary submodular functions. In *Int. Symp. on Mathematical Foundations of Computer Science*, 2009. 1
- [23] T. Werner. A linear programming approach to max-sum problem: A review. *TPAMI*, 29(7), July 2007. 1
- [24] . Yet more details @. <http://www-sop.inria.fr/members/Guillaume.Charpilat/MRF>. 5, 7