Markov Random Field Modeling, Inference & Learning in Computer Vision & Image Understanding: A Survey

Chaohui Wang, Nikos Komodakis, Nikos Paragios

▶ **To cite this version:**

Chaohui Wang, Nikos Komodakis, Nikos Paragios. Markov Random Field Modeling, Inference & Learning in Computer Vision & Image Understanding: A Survey. Computer Vision and Image Understanding, 2013, 117 (11), pp.Page 1610-1627. 10.1016/j.cviu.2013.07.004 . hal-00858390v1

HAL Id: hal-00858390

https://hal.science/hal-00858390v1

Submitted on 5 Sep 2013 (v1), last revised 6 Sep 2013 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Markov Random Field Modeling, Inference & Learning in Computer Vision & Image Understanding: A Survey

Chaohui Wang[a,b], Nikos Komodakis[c], Nikos Paragios[a,d]

[a]*Center for Visual Computing, Ecole Centrale Paris, Grande Voie des Vignes, Châtenay-Malabry, France*
[b]*Perceiving Systems Department, Max Planck Institute for Intelligent Systems, Tübingen, Germany*
[c]*LIGM laboratory, University Paris-East & Ecole des Ponts Paris-Tech, Marne-la-Vallée, France*
[d]*GALEN Group, INRIA Saclay - Île de France, Orsay, France*

Abstract

In this paper, we present a comprehensive survey of Markov Random Fields (MRFs) in computer vision and image understanding, with respect to the modeling, the inference and the learning. While MRFs were introduced into the computer vision field about two decades ago, they started to become a ubiquitous tool for solving visual perception problems around the turn of the millennium following the emergence of efficient inference methods. During the past decade, a variety of MRF models as well as inference and learning methods have been developed for addressing numerous low, mid and high-level vision problems. While most of the literature concerns pairwise MRFs, in recent years we have also witnessed significant progress in higher-order MRFs, which substantially enhances the expressiveness of graph-based models and expands the domain of solvable problems. This survey provides a compact and informative summary of the major literature in this research topic.

Keywords: Markov Random Fields, Graphical Models, MRFs, MAP Inference, Discrete Optimization, MRF Learning

1. Introduction

The goal of computer vision is to enable the machine to understand the world - often called *visual perception* - through the processing of digital signals. Such an understanding for the machine is done by extracting useful information from the digital signals and performing complex reasoning. Mathematically, let **D** denote the observed

data and $\mathbf{x}$ a latent parameter vector that corresponds to a mathematical answer to the visual perception problem. Visual perception can then be formulated as finding a mapping from $\mathbf{D}$ to $\mathbf{x}$, which is essentially an *inverse problem* [1]. Mathematical methods usually model such a mapping through an optimization problem as follows:

$$\mathbf{x}^{\text{opt}} = \arg\min_{\mathbf{x}} E(\mathbf{x}, \mathbf{D}; \mathbf{w}) \ , \tag{1}$$

where the energy (or cost, objective) function $E(\mathbf{x}, \mathbf{D}; \mathbf{w})$ can be regarded as a quality measure of a parameter configuration $\mathbf{x}$ in the solution space given the observed data $\mathbf{D}$, and $\mathbf{w}$ denotes the model parameters[1]. Hence, visual perception involves three main tasks: *modeling*, *inference* and *learning*. The modeling has to accomplish: (i) the choice of an appropriate representation of the solution using a tuple of variables $\mathbf{x}$; and (ii) the design of the class of energy functions $E(\mathbf{x}, \mathbf{D}; \mathbf{w})$ which can correctly measure the connection between $\mathbf{x}$ and $\mathbf{D}$. The inference has to search for the configuration of $\mathbf{x}$ leading to the optimum of the energy function, which corresponds to the solution of the original problem. The learning aims to select the optimal model parameters $\mathbf{w}$ based on the training data.

The main difficulty in the modeling lies in the fact that most of the vision problems are inverse, ill-posed and require a large number of latent and/or observed variables to express the expected variations of the perception answer. Furthermore, the observed signals are usually noisy, incomplete and often only provide a partial view of the desired space. Hence, a successful model usually requires a reasonable *regularization*, a robust *data measure*, and a compact *structure* between the variables of interest to adequately characterize their relationship (which is usually unknown). In the Bayesian paradigm, the *model prior*, the *data likelihood* and the *dependence properties* correspond respectively to these terms, and the maximization of the posterior probability of the latent variables corresponds to the minimization of the energy function in Eq. 1. In addition to these, another issue that should be taken into account during the model-

[1]For the purpose of conciseness, $\mathbf{D}$ and/or $\mathbf{w}$ may not be explicitly written in the energy function in the following presentation unless it is necessary to do so.

ing is the tractability of the inference task, in terms of computational complexity and optimality quality, which introduces additional constraints on the modeling step.

Probabilistic graphical models (usually referred to as *graphical models*) combine probability theory and graph theory towards a natural and powerful formalism for modeling and solving inference and estimation problems in various scientific and engineering fields. In particular, one important type of graphical models - Markov Random Fields (MRFs) - has become a ubiquitous methodology for solving visual perception problems, in terms of both the expressive potential of the modeling process and the optimality properties of the corresponding inference algorithm, due to their ability to model soft contextual constraints between variables and the enormous development of inference methods for such models. Generally speaking, MRFs have the following major useful properties that one can benefit from during the algorithm design. First, MRFs provide a modular, flexible and principled way to combine regularization (or prior), data likelihood terms and other useful cues within a single graph-formulation, where continuous and discrete variables can be simultaneously considered. Second, the graph theoretic side of MRFs provides a simple way to visualize the structure of a model and facilitates the choice and the design of the model. Third, the factorization of the joint probability over a graph could lead to inference problems that can be solved in a computationally efficient manner. In particular, development of inference methods based on discrete optimization enhances the potential of discrete MRFs and significantly enlarges the set of visual perception problems to which MRFs can be applied. Last but not least, the probabilistic side of MRFs gives rise to potential advantages in terms of parameter learning (*e.g.*, [2, 3, 4, 5]) and uncertainty analysis (*e.g.*, [6, 7]) over classic variational methods [8, 9], due to the introduction of probabilistic explanation to the solution [1]. The aforementioned strengths have resulted in the heavy adoption of MRFs towards solving many computer vision, computer graphics and medical imaging problems. During the past decade, different MRF models as well as efficient inference and learning methods have been developed for addressing numerous low, mid and high-level vision problems. While most of the literature is on pairwise MRFs, we have also witnessed significant progress of higher-order MRFs during the recent years, which substantially enhances the expressiveness of graph-based models and expands

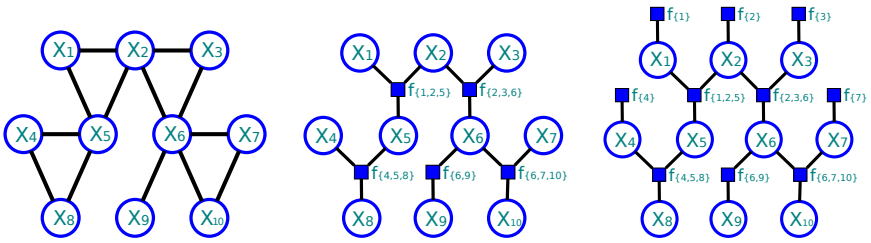
the domain of solvable problems. We believe that a compact and informative summary of the major literature in this research topic will be valuable for the reader to rapidly obtain a global view and hence better understanding of such an important tool.

To this end, we present in this paper a comprehensive survey of MRFs in computer vision and image understanding, with respect to the modeling, the inference and the learning. The remainder of this paper is organized as follows. Section 2 introduces preliminary knowledge on graphical models. In section 3, different important subclasses of MRFs as well as their important applications in visual perception are discussed. Representative techniques for MAP inference in discrete MRFs are presented in section 4. MRF learning techniques are discussed in section 5. Finally, we conclude the survey in section 6.

2. Preliminaries

A graphical model consists of a graph where each node is associated with a random variable and an edge between a pair of nodes encodes probabilistic interaction between the corresponding variables. Each of such models provides a compact representation for a family of joint probability distributions which satisfy the conditional independence properties determined by the topology/structure of the graph: the associated family of joint probability distributions can be factorized into a product of local functions each involving a (usually small) subset of variables. Such a factorization is the key idea of graphical models.

There are two common types of graphical models: *Bayesian Networks* (also known as *Directed Graphical Models* or *Belief Networks*) and *Markov Random Fields* (also known as *Undirected Graphical Models* or *Markov Networks*), corresponding to directed and undirected graphs, respectively. They are used to model different families of distributions with different kinds of conditional independences. It is usually convenient to covert both of them into a unified representation which is called *Factor Graph*, in particular for better visualizing potential functions and performing inference in higher-order models. As preliminaries for the survey, we will proceed with a brief presentation on Markov random fields and factor graphs in the remainder of this sec-



(a) (b) (c)

Figure 1: Examples of Markov Random Fields and Factor Graphs. Note that the Markov random field in (a) can be represented by the two factor graphs (b) and (c). Nevertheless, the factor graph in (c) contains factors corresponding to non-maximal cliques.

tion. We suggest the reader being interested in a larger and more in depth overview the following publications [10, 11, 12, 13].

2.1. Notations

Let us introduce the necessary notations that will be used throughout this survey. For a graphical model, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote the corresponding graph which consists of a set $\mathcal{V}$ of nodes and a set $\mathcal{E}$ of edges. Then, for each node i ($i \in \mathcal{V}$), let X_i denote the associated random variable, x_i the realization of X_i, and $\mathcal{X}_i$ the state space of x_i (*i.e.*, $x_i \in \mathcal{X}_i$). Also, let $\mathbf{X} = (X_i)_{i \in \mathcal{V}}$ denote the joint random variable and $\mathbf{x} = (x_i)_{i \in \mathcal{V}}$ the realization (configuration) of the graphical model taking values in its space $\mathcal{X}$ which is defined as the Cartesian product of the spaces of all individual variables, *i.e.*, $\mathcal{X} = \prod_{i \in \mathcal{V}} \mathcal{X}_i$.

For the purposes of simplification and concreteness, "probability distribution" is used to refer to "probability mass function" (with respect to the counting measure) in discrete cases and "probability density function" (with respect to the Lebesgue measure) in continuous cases. Furthermore, we use $p(x)$ to denote the probability distribution on a random variable X, and use x_c ($c \subseteq \mathcal{V}$) as the shorthand for a tuple c of variables, *i.e.*, $x_c = (x_i)_{i \in c}$. Due to the one-to-one mapping between a node and the associated random variable, we often use "node" to refer to the corresponding random variable in case there is no ambiguity.

2.2. Markov Random Fields (Undirected Graphical Models)

A *Markov Random Field (MRF)* has the structure of an undirected graph $\mathcal{G}$, where all edges of $\mathcal{E}$ are undirected (*e.g.*, Fig. 1(a)), and holds the following local independence assumptions (referred to as *local Markov property*) which impose that a node is independent of any other node given all its neighbors:

$$\forall\, i \in \mathcal{V}, X_i \perp X_{\mathcal{V}-\{i\}} | X_{\mathcal{N}_i} \ , \tag{2}$$

where $\mathcal{N}_i = \{j | \{i, j\} \in \mathcal{E}\}$ denotes the set of neighbors of node i in the graph $\mathcal{G}$, and $X_i \perp X_j | X_k$ denotes the statement that X_i and X_j are independent given X_k. An important notion in MRFs is *clique*, which is defined as a fully connected subset of nodes in the graph. A clique is *maximal* if it is not contained within any other larger clique. The associated family of joint probability distributions are those satisfying the local Markov property (*i.e.*, Eq. 2). According to Hammersley-Clifford theorem [14, 15], they are *Gibbs distributions* which can be factorized into the following form:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{c \in C} \psi_c(x_c) \ , \tag{3}$$

where Z is the normalizing factor (also known as the *partition function*), $\psi_c(x_c)$ denotes the *potential function* of a clique c (or: *clique potential*) which is a positive real-valued function on the possible configuration x_c of the clique c, and C denotes a set of cliques[2] contained in the graph $\mathcal{G}$. We can also verify that any distribution with the factorized form in Eq. 3 satisfies the local Markov property in Eq. 2.

The *global Markov property* consists of all the conditional independences implied within the structure of MRFs, which are defined as: $\forall\, \mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3 \subseteq \mathcal{V}$, if any path from a node in $\mathcal{V}_1$ to a node in $\mathcal{V}_2$ includes at least one node in $\mathcal{V}_3$, then $X_{\mathcal{V}_1} \perp X_{\mathcal{V}_2} | X_{\mathcal{V}_3}$.

[2]Note that any quantities defined on a non-maximal clique can always be redefined on the corresponding maximal clique, and thus C can also consist of only the maximal cliques. However, using only maximal clique potentials may obscure the structure of original cliques by fusing together the potentials defined on a number of non-maximal cliques into a larger clique potential. Compared with such a maximal representation, a non-maximal representation clarifies specific features of the factorization and often can lead to computational efficiency in practice. Hence, without loss of generality, we do not assume that C consists of only maximal cliques in this survey.

Let $\mathcal{I}(\mathcal{G})$ denote the set of such conditional independences. The identification of these independences boils down to a "reachability" problem in graph theory: considering a graph $\mathcal{G}'$ which is obtained by removing the nodes in $\mathcal{V}_3$ as well as the edges connected to these nodes from $\mathcal{G}$, $X_{\mathcal{V}_1} \perp X_{\mathcal{V}_2} | X_{\mathcal{V}_3}$ is true if and only if there is no path in $\mathcal{G}'$ that connects any node in $\mathcal{V}_1 \setminus \mathcal{V}_3$ and any node in $\mathcal{V}_2 \setminus \mathcal{V}_3$. This problem can be solved using standard search algorithms such as breadth-first search (BFS) [16]. Note that the local Markov property and the global Markov property are equivalent for any positive distribution. Hence, if a positive distribution can be factorized into the form in Eq. 3 according to $\mathcal{G}$, then it satisfies all the conditional independences in $\mathcal{I}(\mathcal{G})$. Nevertheless, an distribution instance that can be factorized over $\mathcal{G}$, may satisfy more independences than those in $\mathcal{I}(\mathcal{G})$ [13].

MRFs provide a principled probabilistic framework to model vision problems, thanks to their ability to model soft contextual constraints between random variables [17, 18]. The adoption of such constraints is important in vision problems, since the image and/or scene modeling usually involves interactions between a subset of pixels and/or scene components. Often, these constraints are referred to as "prior" of the whole system. Through MRFs, one can use nodes to model variables of interest and combine different available cues that can be encoded by clique potentials within a unified probabilistic formulation. Then the inference can be performed via *Maximum a posteriori* (MAP) estimation:

$$\mathbf{x}^{\text{opt}} = \arg\max_{\mathbf{x} \in X} p(\mathbf{x}) \ . \tag{4}$$

Since the potential functions are positive, we can define *clique energy* θ_c as a real function on a clique c ($c \in C$):

$$\theta_c(x_c) = -\log \psi_c(x_c) \ . \tag{5}$$

Due to the one-to-one mapping between θ_c and ψ_c, we also refer to θ_c as *potential function* (or *clique potential*) on clique c in the remainder of this survey, leading to a

more convenient representation of the joint distribution $p(\mathbf{x})$:

$$p(\mathbf{x}) = \frac{1}{Z} \exp\{-E(\mathbf{x})\} \ , \tag{6}$$

where $E(\mathbf{x})$ denotes the *energy* of the MRF and is defined as a sum of clique potentials:

$$E(\mathbf{x}) = \sum_{c \in C} \theta_c(x_c) \ . \tag{7}$$

Since the "-log" transformation between the distribution $p(\mathbf{x})$ and the energy $E(\mathbf{x})$ is a monotonic function, the MAP inference in MRFs (Eq. 4) is equivalent to the minimization of $E(\mathbf{x})$ as follows:

$$\mathbf{x}^{\text{opt}} = \arg\min_{\mathbf{x} \in \mathcal{X}} E(\mathbf{x}) \ . \tag{8}$$

In cases of *discrete MRFs* where the random variables are discrete[3] (*i.e.*, $\forall\, i \in \mathcal{V}$, $\mathcal{X}_i$ consists of a discrete set), the above optimization becomes a discrete optimization problem. Numerous works have been done to develop efficient MRF inference algorithms using discrete optimization theories and techniques (*e.g.*, [23, 24, 25, 26, 27, 28, 29, 30, 31]), which have been successfully employed to efficiently solve many vision problems using MRF-based methods (*e.g.*, [32, 33, 34, 35, 36]). Due to the advantages regarding both the modeling and the inference, as discussed previously, discrete MRFs have been widely employed to solve vision problems. We will provide a detailed survey on an important number of representative MRF-based vision models in section 3 and MAP inference methods in section 4.

2.3. Factor Graphs

Factor graph [37, 38] is a unified representation for both BNs and MRFs, which uses additional nodes, named *factor nodes*[4], to explicitly describe the factorization

[3] We should note that *continuous MRFs* have also been used in the literature (*e.g.*, [19, 20, 21]). An important subset of continuous MRFs that has been well studied is *Gaussian MRFs* [22].

[4] We call the nodes in original graphs *usual nodes* when an explicit distinction between the two types of nodes is required to avoid ambiguities.

of the joint distribution in the graph. More specifically, a set $\mathcal{F}$ of factor nodes are introduced into the graph, each corresponding to an objective function term defined on a subset of usual nodes. Each factor encodes a potential function defined on a clique in cases of MRFs[5] (see Eq. 3 or 7). The associated joint probability is a product of factors:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{f \in \mathcal{F}} \phi_f(x_f) \ . \tag{9}$$

Similar to MRFs, we can define the energy of the factor graph as:

$$E(\mathbf{x}) = \sum_{f \in \mathcal{F}} \theta_f(x_f) \ , \tag{10}$$

where $\theta_f(x_f) = -\log \phi_f(x_f)$. Note that there can be more than one factor graphs corresponding to a BN or MRF. Fig. 1(b-c) shows two examples of factor graphs which provide two different possible representations for the MRF in Fig. 1(a).

Factor graphs are bipartite, since there are two types of nodes and no edge exists between two nodes of same types. Such a representation conceptualizes in a clear manner the underlying factorization of the distribution in the graphical model. In particular for MRFs, factor graphs provide a feasible representation to describe explicitly the cliques and the corresponding potential functions when non-maximal cliques are also considered (*e.g.*, Fig. 1(c)). The same objective can be hardly met using the usual graphical representation of MRFs. Computational inference is another strength of factor graphs representations. The *sum-product* and *min-sum* (or: *max-product*[6]) algorithms in the factor graph [38, 11] generalize the classic counterparts [39, 40] in the sense that the order of factors can be greater than two (see Algorithm 1). Furthermore, since an MRF with loops may have no loop in its corresponding factor graph (*e.g.*, see the MRF in Fig. 1(a) and the factor graphs in Fig. 1(b-c)), in such cases the *min-sum* algorithm in the factor graph can perform the MAP inference exactly with polynomial complexity.

[5] Each factor encodes a local conditional probability distribution defined on a usual node and its parents in cases of BNs.

[6] The *max-product* algorithm is to maximize the probability $p(\mathbf{x})$ which is a product of local functions (Eq. 9), while the *min-sum* algorithm is to minimize the corresponding energy which is a sum of local energy functions (Eq. 10). They are essentially the same algorithm.

Such factor graphs without loop (*e.g.*, Fig. 1(b-c)) are referred to as *factor trees*.

3. MRF-based Vision Models

According to the order of interactions between variables, MRF models can be classified into *pairwise models* and *higher-order models*. Another important class is *Conditional Random Fields* (CRFs). Below, we present these three typical models that are commonly used in vision community.

3.1. Pairwise MRF Models

The most common type of MRFs that is widely used in computer vision is the *pairwise MRF*, in which the associated energy is factorized into a sum of potential functions defined on cliques of order strictly less than three. More specifically, a pairwise MRF consists of a graph $\mathcal{G}$ with a set $(\theta_i(\cdot))_{i\in\mathcal{V}}$ of *unary potentials* (also called *singleton potentials*) defined on single variables and a set $(\theta_{ij}(\cdot))_{\{i,j\}\in\mathcal{E}}$ of *pairwise potentials* defined on pairs of variables. The MRF energy has the following form:

$$E(\mathbf{x}) = \sum_{i\in\mathcal{V}} \theta_i(x_i) + \sum_{\{i,j\}\in\mathcal{E}} \theta_{ij}(x_{ij}) \ . \tag{11}$$

Pairwise MRFs have attracted the attention of a lot of researchers and numerous works have been done in past decades, mainly due to the facts that pairwise MRFs inherit simplicity and computational efficiency, and that the interaction between pairs of variables is the most common and fundamental type of interactions required to model many vision problems. In computer vision, such works include both the modeling of vision problems using pairwise MRFs (*e.g.*, [41, 42, 43, 36, 44]) and the efficient inference in pairwise MRFs (*e.g.*, [23, 26, 28, 27, 45]). Two most typical graph structures used in computer vision are *grid-like structures* (*e.g.*, Fig. 2) and *part-based structures* (*e.g.*, Fig. 3). Grid-like structures provide a natural and reasonable representation for images, while part-based structures are often associated with deformable and/or articulated objects.

(a) 4-neighborhood system (b) 8-neighborhood system

Figure 2: Examples of MRFs with Grid-like Structures

3.1.1. Grid-like Models

Pairwise MRFs of *grid-like structures* (Fig. 2) have been widely used in computer vision to deal with numerous important problems, such as image denoising/restoration (*e.g.*, [41, 46, 47]), super-resolution (*e.g.*, [48, 49, 50]), stereo vision/multi-view reconstruction (*e.g.*, [51, 32, 52]), optical flow and motion analysis (*e.g.*, [53, 54, 55, 56]), image registration and matching (*e.g.*, [33, 57, 58, 59]), segmentation (*e.g.*, [60, 42, 36, 61]) and over-segmentation (*e.g.*, [62, 63, 64]).

In this context, the nodes of an MRF correspond to the lattice of pixels[7]. The edges corresponding to pairs of neighbor nodes are considered to encode contextual constraints between nodes. The random variable x_i associated with each node i represents a physical quantity specific to problems[8] (*e.g.*, an index denoting the segment to which the corresponding pixel belongs for image segmentation problem, an integral value between 0 and 255 denoting the intensity of the corresponding pixel for gray image denoising problem, *etc*.). The data likelihood is encoded by the sum of the unary potentials $\theta_i(\cdot)$, whose definition is specific to the considered application (*e.g.*, for image denoising, such unary terms are often defined as a penalty function based on the deviation of the observed value from the underlying value). The contextual constraints compose a prior model on the configuration of the MRF, which is usually encoded by the sum of all the pairwise potentials $\theta_{ij}(\cdot,\cdot)$. The most typical and commonly used contextual constraint is the *smoothness*, which imposes that physical quantities corresponding to the states of nodes varies "smoothly" in the spatial domain as defined by

[7] Other homogeneously distributed units such as 3D voxels and control points [33] can also be considered in such MRFs.

[8] An MRF is called *binary MRF* if each node has only two possible values, 0 or 1.

the connectivity of the graph. To this end, the pairwise potential $\theta_{ij}(\cdot,\cdot)$ between a pair $\{i, j\}$ of neighbor nodes is defined as a cost term that penalizes the variation of the states between the two nodes:

$$\theta_{ij}(x_{ij}) = \rho(x_i - x_j) \ , \tag{12}$$

where $\rho(\cdot)$ is usually an even and non-decreasing function. In computer vision, common choices for $\rho(\cdot)$ are *(generalized) Potts model*[9] [66, 67], *truncated absolute distance* and *truncated quadratic*, which are typical *discontinuity preserving* penalties:

$$\rho(x_i - x_j) = \begin{cases} w_{ij} \cdot (1 - \delta(x_i - x_j)) & \text{(Potts models)} \\ \min(K_{ij}, |x_i - x_j|) & \text{(truncated absolute distance)} \\ \min(K_{ij}, (x_i - x_j)^2) & \text{(truncated quadratic)} \end{cases} , \tag{13}$$

where $w_{ij} \geq 0$ is a weight coefficient[10] for the penalties, Kronecker delta $\delta(x)$ is equal to 1 when $x = 0$, and 0 otherwise, and K_{ij} is a coefficient representing the maximum penalty allowed in the truncated models. More discontinuity preserving regularization functions can be found in for example [68, 69]. Last, it should be mentioned that pairwise potentials in such grid-like MRFs can also be used to encode other contextual constraints, such as *star shape priors* [70], *compact shape priors* [71], *layer constraints* [62], *Hausdorff distance priors* [72] and *ordering constraints* [73, 74].

The grid-like MRF presented above can be naturally extended from pixels to other units. For example, there exist works that use superpixel primitives instead of pixel primitives when dealing with images (*e.g.*, [75, 76]), mainly aiming to gain computational efficiency and/or use superpixels as regions of support to compute features for other mid-level and high-level vision applications. Another important case is the segmentation, registration and tracking of 3D surface meshes (*e.g.*, [77, 78]), where we aim to infer the configuration of each vertex or facet on the surface. In these cases, the node of MRFs can be used to model the superpixel, vertex or face, nevertheless, the topology could be a less regular grid.

[9] Note that *Ising model* [65, 41] is a particular case of *Potts model* where each node has two possible states.

[10] w_{ij} is a constant for all pairs $\{i, j\}$ of nodes in the original Potts model in [66].

(a) Pictorial Structure (b) MRF model corresponding to (a)

Figure 3: Example of MRFs with Pictorial Structures (The original image used in (a) is from *HumanEva-I* database [79]: `http://vision.cs.brown.edu/humaneva/`.)

3.1.2. Part-based Models

MRFs of *pictorial structures* (Fig. 3) provide a natural part-based modeling tool for representing deformable objects and in particular articulated objects. Their nodes correspond to components of such objects. The corresponding latent variables represent the spatial pose of the components. An edge between a pair of nodes encode various interactions such as kinematic constraints between the corresponding pair of components. In [43], *Pictorial model* [80] was employed to deal with pose recognition of human body and face efficiently with dynamic programming. In this work, a tree-like MRF (see Fig. 3) was employed to model spring-like priors between pairs of components through pairwise potentials, while the data likelihood is encoded in the unary potentials each of which is computed from the appearance model of the corresponding component. The pose parameters of all the components are estimated though the MAP inference, which can be done very efficiently in such a tree-structured MRF using dynamic programming [81, 16] (*i.e.*, min-sum belief propagation [39, 40, 11]).

Later, part-based models have been adopted and/or extended to deal with the pose estimation, detection and tracking of deformable object such as human body [20, 82, 83, 84, 85], hand [86, 87] and other objects [88, 89]. In [88], the part-based model was extended, with respect to that of [43], regarding the topology of the MRF as well as the image likelihood in order to deal with the pose estimation of animals such as cows and horses. The topology of part-based models was also extend to other typical graphs such as *k-fans graphs* [90, 91] and *out-planer graphs* [92]. Pictorial structures conditioned on *poselets* [93] were proposed in [85] to incorporate higher-order dependency between the parts of the model while keeping the inference efficient (since the model becomes tree-structured at the graph-inference stage). Continuous MRFs of pic-

torial structures were proposed in [20] and [86] to deal with body and/or hand tracking, where nonparametric belief propagation algorithms [19, 21] were employed to perform inference. In the subsequent papers [82, 87], occlusion reasoning was introduced into their graphical models in order to deal with occlusions between different components. Indeed, the wide existence of such occlusions in the cases of articulated objects is an important limitation of the part-based modeling. Recently, a rigorous visibility modeling in graphical models was achieved in [94] via the proposed joint 2.5D layered model where top-down scene-level and bottom-up pixel-level representations are seamlessly combined through local constraints that involve only pairs of variables (as opposed to previous 2.5D layered models where the depth ordering was commonly modeled as a total and strict order between all the objects), based on which image segmentation (pixel-level task), multi-object tracking and depth ordering (scene-level tasks) are simultaneously performed via a single pairwise MRF model.

The notion of "part" can also refer to a feature point or landmark distributed on the surface of an object. In such a case, MRFs provide a powerful tool for modeling prior knowledge (*e.g.*, generality and intra-class variations) on a class of shapes, which is referred to as *statistical shape modeling* [95]. The characterization of shape priors using local interactions (*e.g.*, statistics on the Euclidean distance) between points can lead to useful properties such as translation and rotation invariances with respect to the global pose of the object in the observed image. Together with efficient inference methods, such MRF-based prior models have been employed to efficiently solving problems related to the inference of the shape model such as knowledge-based object segmentation (*e.g.*, [96, 97]). However, the factorization of probability or energy terms into an MRF can be very challenging, where good approximate solutions may be resorted to (*e.g.*, [97, 98]). In this line of research, recently [99] proposed to employ *divergence theorem* to exactly factorize regional data likelihood in their pairwise MRF model for object segmentation.

Remark

The computer vision community has primarily focused on pairwise MRF models where interactions between parameters were often at the level of pairs of variables.

This was a convenient approach driven mostly from the optimization viewpoint since pairwise MRFs inherit the lowest rank of interactions between variables and numerous efficient algorithms exist for performing inference in such models. Such interactions to certain extent can cope with numerous vision problems (segmentation, estimation, motion analysis and object tracking, disparity estimation from calibrated views, *etc*.). However, their limitations manifest when a better performance is desired for those problems or when graph-based solutions are resorted to for solving more complex vision problems, where higher-order interactions between variables are needed to be modeled. One the other hand, the rapid development of computer hardwares in terms of memory capacity and CPU speed provides the practical base and motivates the consideration of higher-order interactions in vision models. In such a context, higher-order MRF models has attracted more and more attentions, and many related vision models and inference methods have been proposed.

3.2. Higher-order MRF Models

Higher-order MRFs[11] involve potential functions that are defined on cliques containing more than two nodes and cannot be further decomposed. Such higher-order potentials, compared to pairwise ones, allow a better characterization of statistics between random variables and increase largely the ability of graph-based modeling. We summary below three main explorations of such advantages in solving vision problems.

First, for many vision problems that already were addressed by pairwise models, higher-order MRFs are often adopted to model more complex and/or natural statistics as well as richer interactions between random variables, in order to improve the performance of the method. One can cite for example the higher-order MRF model proposed in [100, 101] to better characterize image priors, by using the Product-of-Experts framework to define the higher-order potentials. Such a higher-order model was successfully applied in image denoising and inpainting problems [100, 101]. $\mathcal{P}^n$ *Potts model* was proposed in [102, 103], which considers a similar interaction as the generalized Potts model [67] (see Eq. 13), but between n nodes instead of between

[11] They are also referred to as *high-order MRFs* in part of the literature.

two nodes, and leads to better performance in image segmentation. This model is a strict generalization of the generalized Potts model and has been further enriched towards *robust* $\mathcal{P}^n$ *model* in [104, 105]. [106] used higher-order smoothness priors for addressing stereo reconstruction problems, leading better performance than pairwise smoothness priors. Other types of higher-order pattern potentials were also considered in [107] to deal with image/signal denoising and image segmentation problems. All these works demonstrated that the inclusion of higher-order interactions is able to significantly improve the performance compared to pairwise models in the considered vision problems.

Higher-order models become even more important in cases where we need to model measures that intrinsically involve more than two variables. A simple example is the modeling of second-order derivative (or even higher-order derivatives), which is often used to measure bending force in shape prior modeling such as active contour models (*i.e.*, "Snake") [108]. In [109], dynamic programming was adopted to solve "Snake" model in a discrete setting, which is essentially a higher-order MRF model. A third-order spatial priors based on second derivatives was also introduced to deal with image registration in [110]. In the optical flow formulation proposed in [111], higher-order potentials were used to encode angle deviation prior, non-affine motion prior as well as the data likelihood. [112] proposed a compact higher-order model that encodes a curvature prior for pixel labeling problem and demonstrated its performance in image segmentation and shape inpainting problems. *Box priors* were introduced in [113] for performing image segmentation given a user-provided object bounding box, where topological constraints defined based on the bounding box are incorporated into the whole optimization formulation and have been demonstrated to be able to prevent the segmentation result from over-shrinking and ensure the tightness of the object boundary delimited by the user-provided box. [114] proposed a higher-order illumination model to couple the illumination, the scene and the image together so as to jointly recover the illumination environment, scene parameters, and an estimate of the cast shadows given a single image and coarse initial 3D geometry. Another important motivation for employing higher-order models is to characterize statistics that are invariant with respect to global transformation when dealing with deformable shape inference

[115, 116]. Such approaches avoid explicit estimation of the global transformation such as 3D pose (translation, rotation and scaling) and/or camera viewpoint, which is substantially beneficial to both the learning and the inference of the shape model.

Meanwhile, *global models*, which include potentials involving all the nodes, have been developed, together with the inference algorithms for them. For example, global connectivity priors (*e.g.*, the foreground segment must be connected) were used in [117] and [118] to enforce the connectedness of the resulting pixel labeling in binary image segmentation, which were shown to be able to achieve better performance compared to merely using Potts-model with smoothness terms (see section 3.1.1). In order to deal with unsupervised image segmentation where the number of segments are unknown in advance, [119, 120] introduced 'label costs" [121] into graph-based segmentation formulation, which imposes a penalty to a label l (or a subset $\mathcal{L}_s$ of labels) from the predefined possible label set $\mathcal{L}$ if at least one node is labeled as l (or an element in $\mathcal{L}_s$) in the final labeling result. By doing so, the algorithm automatically determines a subset of labels from $\mathcal{L}$ that are finally used, which corresponds to a model selection process. Another work in a similar line of research is presented in [122, 123], where "object co-occurrence statistics" - a measure of which labels are likely to appear together in the labeling result - are incorporated within traditional pairwise MRF/CRF models for addressing object class image segmentation and have been shown to improve significantly the segmentation performance.

3.3. Conditional Random Fields

A Conditional Random Field (CRF) [124, 125] encodes, with the same concept as the MRF earlier described, a conditional distribution $p(\mathbf{X}|\mathbf{D})$ where $\mathbf{X}$ denotes a tuple of latent variables and $\mathbf{D}$ a tuple of observed variables (data). Accordingly, the Markov properties for the CRF are defined on the conditional distribution $p(\mathbf{X}|\mathbf{D})$. The local Markov properties in such a context become:

$$\forall\, i \in \mathcal{V}, X_i \perp X_{\mathcal{V}-\{i\}} | \{X_{N_i}, \mathbf{D}\} \ , \tag{14}$$

while the global Markov property can also be defined accordingly. The conditional distribution $p(\mathbf{X}|\mathbf{D})$ over the latent variables $\mathbf{X}$ is also a Gibbs distribution and can be written as the following form:

$$p(\mathbf{x}|\mathbf{D}) = \frac{1}{Z(\mathbf{D})} \exp\{-E(\mathbf{x};\mathbf{D})\} \ , \tag{15}$$

where the energy $E(\mathbf{x};\mathbf{D})$ of the CRF is defined as:

$$E(\mathbf{x};\mathbf{D}) = \sum_{c \in C} \theta_c(x_c;\mathbf{D}) \ . \tag{16}$$

We can observe that there is no modeling on the probabilistic distribution over the variable in $\mathbf{D}$, which relaxes the concern on the dependencies between these observed variables, whereas such dependencies can be rather complex. Hence, CRFs reduce significantly difficulty in modeling the joint distribution of the latent and observed variables, and consequently, observed variables can be incorporated into the CRF framework in a more flexible way. Such a flexibility is one of the most important advantages of CRFs compared with generative MRFs[12] when used to model a system. For example, the fact that clique potentials can be data dependent in CRFs could lead to more informative interactions than data independent clique potentials. Such an concept was adopted for example in binary image segmentation [127], where the intensity contrast and the spatial distance between neighbor pixels are employed to modulate the values of pairwise potentials of a grid-like CRF, as opposed to Potts models (see section 3.1.1). Despite the difference in the probabilistic explanation, the MAP inferences in generative MRFs and CRFs boil down to the same problem.

CRFs have been applied to various fields such as computer vision, bioinformatics and text processing among others. In computer vision, besides [127], grid-like CRFs were also employed in [128] to model spatial dependencies in the image, leading to a data-dependent smoothness terms between neighbor pixels. With the learned parameters from training data, a better performance has been achieved in the image restoration

[12] Like [126], we use the term *generative MRFs* to distinguish the usual MRFs from CRFs.

experiments compared to the classic Ising MRF model [41]. Hierarchical CRFs have also been developed to incorporate features from different levels so as to better perform object class image segmentation. One can cite for example the multi-scale CRF model introduced in [129] and "associative hierarchical CRFs" proposed in [130]. Moreover, CRFs have also been applied for object recognition/detection. For example, a discriminative part-based approach was proposed in [131] to recognize objects based on a tree-structured CRF. In [132], object detectors were combined within a CRF model, leading to an efficient algorithm to jointly estimate the class category, location, and segmentation of objects/regions from 2D images. Last, it is worth mentioning that recently, based on a mean field approximation to the CRF distribution, [133] proposed a very efficient approximate inference algorithm for fully connected grid-like CRFs where pairwise potentials corresponds to a linear combination of Gaussian kernels, and demonstrated that such a dense connectivity at the pixel level significantly improves the accuracy in class segmentation compared to 4-neighborhood system (Fig. 2) [134] and *robust* $\mathcal{P}^n$ *model* [105]. Their techniques were further adopted and extended to address optical flow computing [135, 136], and to address cases where pairwise potentials are non-linear dissimilarity measures that do not required to be distance metrics [137].

4. MAP Inference Methods

An essential problem regarding the application of MRF models is how to infer the optimal configuration for each of the nodes. Here, we focus on the MAP inference (*i.e.*, Eq. 4) in discrete MRFs, which boils down to an energy minimization problem as shown in Eq. 8. Such a combinatorial problem is known to be NP-hard in general [23, 25], except for some particular cases such as MRFs of bounded tree-width [138, 139, 12] (*e.g.*, tree-structured MRFs [39]) and pairwise MRFs with submodular energy [25, 140].

The most well-known early (before the 1990s) algorithms for optimizing the MRF energy were *iterated conditional modes* (ICM) [141], *simulated annealing* methods (*e.g.*, [41, 142, 143]) and *highest confidence first (HCF)* [144, 145]. While being computational efficient, ICM and HCF suffer from their ability to recover a good optimum.

On the other hand, for simulated annealing methods, even if in theory they provide certain guarantees on the quality of the obtained solution, in practice from computational viewpoint such methods are impractical. In the 1990s, more advanced methods, such as *loopy belief propagation* (LBP) (*e.g.*, [48, 146, 147]) and *graph cuts* techniques (*e.g.*, [46, 51, 67, 148, 23]), provided powerful alternatives to the aforementioned methods from both computational and theoretical viewpoints and have been used to solve numerous visual perception problems (*e.g.*, [48, 58, 46, 148, 32, 60, 42]). Since then, the MRF optimization is experiencing a renaissance, and more and more researchers have been working on it. For recent MRF optimization techniques, one can cite for example *QPBO* techniques (*e.g.*, [149, 150, 151, 152]), LP primal-dual algorithms (*e.g.*, [153, 154, 29]) as well as dual methods (*e.g.*, [26, 28, 154, 155]).

There exist three main classes of MAP inference methods for pairwise MRFs and they also have been extended to deal with higher-order MRFs. In order to provide an overview of them, in this section we will first review *graph cuts* and their extensions for minimizing the energy of pairwise MRFs in section 4.1. Then in section 4.2 and Appendix B, we will describe the *min-sum belief propagation* algorithm in factor trees and also show its extensions to dealing with an arbitrary pairwise MRF. Following that, we review in section 4.3 recent developed dual methods for pairwise MRFs, such as *tree-reweighted message passing* methods (*e.g.*, [26, 28]) and *dual-decomposition* approaches (*e.g.*, [154, 156]). Last but not least, a survey on MRF inference methods for higher-order MRFs will be provided in section 4.4.

4.1. Graph Cuts and Extensions

Graph cuts consist of a family of discrete algorithms that use *min-cut/max-flow* techniques to efficiently minimize the energy of discrete MRFs and have been used to solve many vision problems (*e.g.*, [46, 148, 42, 32, 36, 34]).

The basic idea of graph cuts is to construct a directed graph $\mathcal{G}^{st} = (\mathcal{V}^{st}, \mathcal{E}^{st})$ (called *s-t graph*[13]) with two special terminal nodes (*i.e.*, the source s and the sink t) and non-negative capacity setting $c(i, j)$ on each directed edge $(i, j) \in \mathcal{E}^{st}$, such that the cost

[13] Note that generations such as *multi-way cut* problem [157] which involves more than two terminal nodes are NP-hard.

$C(S,T)$ (Eq. 17) of the s-t cut that partitions the nodes into two disjoint sets (S and T such that $s \in S$ and $t \in T$) is equal to the energy of the MRF with the corresponding configuration[14] $\mathbf{x}$ (up to a constant difference):

$$C(S,T) = \sum_{i \in S, j \in T, (i,j) \in \mathcal{E}^{st}} c(i,j) \quad . \tag{17}$$

An MRF that has such an s-t graph is called *graph-representable*[15] and can be solved in polynomial time using graph cuts [25]. The minimization of the energy of such an MRF is equivalent to the minimization of the cost of the s-t-cut problem (*i.e.*, min-cut problem). The Ford and Fulkerson theorem [158] states that the solution of the min-cut problem corresponds to the maximum flow from the source s to the sink t (*i.e.*, max-flow problem). Such a problem can be efficiently solved in polynomial time using many existing algorithms such as Ford-Fulkerson style augmenting paths algorithms [158] and Goldberg-Tarjan style push-relabel algorithms [159]. Note that the min-cut problem and the max-flow problem are actually dual LP problems of each other [160].

Unfortunately, not all the MRFs are graph-representable. Previous works have been done to explore the class of graph-representable MRFs (*e.g.*, [161, 24, 25, 140]). They demonstrated that a pairwise discrete MRF is graph-representable so that the global minimum of the energy can be achieved in polynomial time via graph cuts, if the energy function of the MRF is submodular (see Appendix A for the definition of submodularity). However, in numerous vision problems, more challenging energy functions that do not satisfy the submodular condition (Eq. A.1) are often required. The minimization of such non-submodular energy functions is NP-hard in general [23, 25] and an approximation algorithm would be required to approach the global optimum.

More than two decades ago, [46] first proposed to use min-cut/max-flow techniques to exactly optimize the energy of a binary MRF (*i.e.*, Ising model) for image restoration in polynomial time. However, the use of such min-cut/max-flow techniques did

[14] The following rule can be used to associate an s-t cut to an MRF labeling: for a node $i \in \mathcal{V}^{st} - \{s,t\}$, i) if $i \in S$, the label x_i of the corresponding node in the MRF is equal to 0; ii) if $i \in T$, the label x_i of the corresponding node in the MRF is equal to 1.

[15] Note that, in general, such an s-t graph is not unique for a graph-representable MRF.

not draw much attention in computer vision community in the following decade since then, probably due to the fact that the work was published in a journal of statistics community and/or that the model considered in [46] is quite simple. Such a situation has changed in late 1990s when a number of techniques based on graph cuts were proposed to solve more complicated MRFs. One can cite for example the works described in [67, 51, 148], which proposed to use min-cut/max-flow techniques to minimize multi-label MRFs. In particular, the work introduced in [67] achieved, based on the proposed optimization algorithms, much more accurate results than the state-of-the-art in computing stereo depth, and thus motivated the use of their optimization algorithms for many other problems (*e.g.*, [162, 163, 164]), also leading to excellent performance. This significantly popularized graph cuts techniques in computer vision community. Since then, numerous works have been done for exploring larger subsets of MRFs that can be exactly or approximately optimized by graph cuts and for developing more efficient graph-cuts-based algorithms.

Towards Multi-label MRFs

There are two main methodologies for solving multi-label MRFs based on graph cuts: *label-reduction* and *move-making*.

The first methodology (*i.e.*, *label-reduction*) is based on the observation that some solvable types of multi-label MRFs can be exactly solved in polynomial time using graph cuts by first introducing auxiliary binary variables each corresponding to a possible label of a node and then deriving a min-cut problem that is equivalent to the energy minimization of the original MRF. We can cite for example an efficient graph construction method proposed in [24] to deal with arbitrary *convex* pairwise MRFs, which was further extended to submodular pairwise MRFs in [140]. Such a methodology can perform MAP inference in some types of MRFs. However, the solvable types are quite limited, since it is required that the obtained binary MRF (via introducing auxiliary binary variables) should be graph-representable. Whereas, the other optimization methodology (*i.e.*, *move-making*) provides a very important tool for addressing larger sub-classes of MRFs.

The main idea of *move-making* is to optimize the MRF energy by defining a set of

proposals (*i.e.*, possible "moves") based on the initial MRF configuration and choosing the best move as the initial configuration for the next iteration, which is done iteratively until the convergence when no move leads to a lower energy. The performance of an algorithm developed based on such a methodology mainly depends on the size (denoted by $\mathcal{M}$) of the set of proposals at each iteration. For example, ICM [141] iteratively optimizes the MRF energy with respect to a node by fixing the configuration of all the other nodes. It can be regarded as the simplest move-making approach, where $\mathcal{M}$ is equal to the number of labels of the node that is considered to make move at an iteration. ICM has been shown to perform poorly when dealing with vision MRF models, due to the small set $\mathcal{M}$ of proposals [35].

Graph-cuts-based methods have been proposed to exponentially increase the size of the set $\mathcal{M}$ of proposals, for example, by considering the combination of two possible values for all the nodes ($\mathcal{M} = 2^{|\mathcal{V}|}$). In the representative works of [165, 23], *α-expansion* and *$\alpha\beta$-swap* were introduced to generalize binary graph cuts to handle pairwise MRFs with *metric* and/or *semi-metric* energy. An α-expansion refers to a move from $\mathbf{x}$ to $\mathbf{x}'$ such that: $x_i \neq x'_i \Rightarrow x'_i = \alpha$. An $\alpha\beta$-swap means a move from $\mathbf{x}$ to $\mathbf{x}'$ such that: $x_i \neq x'_i \Rightarrow x_i, x'_i \in \{\alpha, \beta\}$. [165, 23] proposed efficient algorithms for determining the optimal expansion or swap moves by converting the problems into binary labeling problems which can be solved efficiently using graph cuts techniques. In such methods, a drastically larger $\mathcal{M}$ compared to that of ICM makes the optimization less prone to be trapped at local minima and thus leads to much better performance [35]. Moreover, unlike ICM which has no optimum quality guarantee, the solution obtained by α-expansion has been proven to possess a bounded ratio between the obtained energy and the global optimal energy [165, 23].

In addition, *range moves* methods [166, 167, 168] have been developed based on min-cut/max-flow techniques to improve the optimum quality in addressing MRFs with truncated convex priors. Such methods explore a large search space by considering a range of labels (*i.e.*, an interval of consecutive labels), instead of dealing with one/two labels at each iteration as what is done in α-expansion or $\alpha\beta$-swap. In particular, range expansion has been demonstrated in [167] to provide the same multiplicative bounds as the standard linear programming (LP) relaxation (see section 4.3) in polynomial time,

and to provide a faster algorithm for dealing with the class of MRFs with truncated convex priors compared to LP-relaxation-based algorithms such as tree-reweighted Message Passing (TRW) techniques (see section 4.3). Very recently, [169] proposed a dynamic-programming-based algorithm for approximately performing α-expansion, which significantly speeds up the original α-expansion algorithm [165, 23].

Last, we should note that expansion is a very important concept in optimizing the energy of a multi-label MRF using graph cuts. Many other works in this direction are based on or partially related to it, which will be reflected in the following discussion.

Towards Non-submodular Functions

Graph cuts techniques have also been extended to deal with non-submodular binary energy functions. *Roof duality* was proposed in [170], which provides an LP relaxation approach to achieving a partial optimal labeling for quadratic pseudo-boolean functions (the solution will be a complete labeling that corresponds to global optimum if the energy is submodular). The *persistency* property of roof duality indicates that the configurations of all the labeled nodes are exactly those corresponding to the global optimum. Hence, QPBO at least provides us with a partial labeling of the MRF and the number of unlabeled nodes depends on the number of nonsubmodular terms included in the MRF. Such a method was efficiently implemented in [149], which is referred to as *Quadratic Pseudo-Boolean Optimization (QPBO)* algorithm and can be regarded as a graph-cuts-based algorithm with a special graph construction where two nodes in s-t graph are used to represent two complementary states of a node in the original MRF [150]. By solving min-cut/max-flow in such an s-t graph, QPBO outputs a solution assigning 0, 1 or $\frac{1}{2}$ to each node in the original MRF, where the label $\frac{1}{2}$ means the corresponding node is *unlabeled*.

Furthermore, two different techniques were introduced in order to extend QPBO towards achieving a complete solution. One is *probing* (called *QPBO-P*) [151, 152], which aims to gradually reduce the number of unlabeled nodes (either by finding the optimal label for certain unlabeled nodes or by regrouping a set of unlabeled nodes) until convergence by iteratively fixing the label of a unlabeled node and performing QPBO. The other one is *improving* (called *QPBO-I*) [152], which starts from a com-

plete labeling $\mathbf{y}$ and gradually improves such a labeling by iteratively fixing the labels of a subset of nodes as those specified $\mathbf{y}$ and using QPBO to get a partial labeling to update $\mathbf{y}$.

Besides, QPBO techniques have been further combined with the *label-reduction* and *move-making* techniques presented previously to deal with multi-label MRFs. For the former case, in [171], a multi-label MRF is converted into an equivalent binary MRF [24] and then QPBO techniques are employed to solve the linear relaxation of the obtained binary MRF. It provides a partial optimal labeling for multi-label MRFs. Nevertheless, a disadvantage of such an approach is the expensive computational complexity. For the latter case, an interesting combination of QPBO and move-making techniques was proposed in [172], which is referred to as *fusion moves*. Given two arbitrary proposals $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ of the full labeling of the MRF, fusion moves combine the proposals together via a binary labeling problem, which is solved using QPBO so as to achieve a new labeling $\mathbf{x}'$ such that: $\forall i, x'_i \in \{x_i^{(1)}, x_i^{(2)}\}$. Using the proposed label selection rule, $\mathbf{x}'$ is guaranteed to have an energy lower or equal than the energies of both proposals $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$. Hence, *fusion moves* provides an effective tool for addressing the optimization of multi-label discrete/continuous MRFs. In addition, it turns out that fusion moves generalize some previous graph-cuts-based methods such as α*-expansion* and $\alpha\beta$*-swap*, in the sense that the latter methods can be formulated as fusion moves with particular choices of proposals. This suggests that fusion moves can serve as building block within various existing optimization schemes so as to develop new techniques, such as the approaches proposed in [172] for the parallelization of MRF optimization into several threads and the optimization of continuous-labeled MRFs with 2D labels.

Towards Improving Efficiency

We should also note that different methods have been developed to increase the efficiency of graph-cuts-based algorithms, in particular in the context of dynamic MRFs (*i.e.*, the potential functions vary over time, whereas the change between two successive instants is usually quite small). Below are several representative works in this line of research.

A dynamic max-flow algorithm (referred to as *dynamic graph cuts*) was proposed in [173, 27] to accelerate graph cuts when dealing with dynamics MRFs, where the key idea is to reuse the flow obtained by solving the previous MRF to initialize the min-cut/max-flow problems so as to significantly reduce the computational time of min-cut. Another dynamic algorithm was also proposed in [174] to improve the convergence of optimization for dynamic MRFs, by using the min-cut solution of the previous MRF to generate an initialization for solving the current MRF.

In [154, 29], a primal-dual scheme based on linear programming relaxation (referred to as *FastPD*) was proposed for optimizing the MRF energy, by recovering pair of solutions for the primal and the dual such that the gap between them is minimized[16]. This method exploits information coming from both the original MRF optimization problem and its dual problem, and achieves a substantial speedup with respect to previous methods such as [23] and [153]. In addition, it can also speed up the optimization in the case of dynamic MRFs, where one should expect that the new pair of primal-dual solutions is closed to the previous one.

Besides, [175, 176] proposed two similar but simpler techniques with respect to that of [154, 29] to achieve a similar computational efficiency. The main idea of the first one (referred to as *dynamic α-expansion*) is to "recycle" results from previous problem instances. Similar to [173, 27, 174], the flow from the corresponding move in the previous iteration is reused for solving an expansion move in a particular iteration. And when dealing with dynamic MRFs, the primal and dual solutions obtained from the previous MRF are used to initialize the min-cut/max-flow problems for the current MRF. The second method aims to simplify the energy function by solving partial optimal MRF labeling problems [171, 177] and reducing the number of unlabeled variables, while the dual (flow) solutions of such problems are used to generate a "good" initialization for the dynamic α-expansion algorithm.

Last but not least, based on the primal-dual interpretation of the expansion algorithm introduced by [154, 29], [178] proposed an approach to optimize the choice of the move space for each iteration by exploiting the primal-dual gap. As opposed to

[16] *FastPD* can also be viewed as a generalization of α-expansion.

traditional move-making methods that search for better solutions in some pre-defined moves spaces around the current solution, such an approach aims to greedily determine the move-space (*e.g.*, the optimal value of α in the context of α-expansion) that will lead to largest decrease in the primal-dual gap at each iteration. It was demonstrated experimentally to increase significantly the optimization efficiency.

4.2. Belief Propagation Algorithms

Belief propagation algorithms use local message passing to perform inference on graphical models. They provide an exact inference algorithm for tree-structured discrete MRFs, while an approximate solution can be achieved for a loopy graph. In particular, for those loopy graphs with low tree-widths such as cycles, extended belief propagation methods such as *junction tree algorithm* [138, 139, 12] provide an efficient algorithm to perform exact inference. These belief propagation algorithms have been adopted to perform MAP inference in MRF models for a variety of vision problems (*e.g.*, [43, 48, 58, 179, 92]).

4.2.1. Belief Propagation in Tree

Belief propagation (BP) [39, 40, 11] was proposed originally for exactly solving MAP inference (*min-sum* algorithm) and/or maximum-marginal inference (*sum-product* algorithm) in a tree-structured graphical model in polynomial time. This type of methods can be viewed as a special case of *dynamic programming* in graphical models [81, 16, 180]. A representative vision model that can be efficiently solved by BP is the pictorial model [80, 43] (see section 3.1.2).

In the *min-sum* algorithm[17] for a tree-structured MRF, a particular node is usually designated as the "root" of the tree. Then messages are propagated inwards from the leaves of the tree towards the root, where each node sends its message to its parent once it has received all incoming messages from its children. During the message passing, a local lookup table is generated for each node, recording the optimal labels of

[17] Note that all the BP-based algorithms presented in section 4.2 include both *min-sum* and *sum-product* versions. We focus here on the *min-sum* version. Nevertheless, the *sum-product* version can be easily obtained by replacing the message computation with the sum of the product of function terms. We refer the reader to [38, 11, 12] for more details.

all children for each of its possible labels. Once all messages arrive at the root node, a minimization is performed over the sum of the messages and the unary potentials of the root node, giving the minimum value for the MRF energy as well as the optimal label for the root node. In order to determine the labels for the other nodes, the optimal label is then propagated outwards from the root to the leaves of the tree, simply via checking the lookup tables obtained previously, which is usually referred to as *back-tracking*. A detailed algorithm is provided in Algorithm 1 (Appendix B) based on the factor graph representation [38, 11], since as we mentioned in section 2.3, the factor graph makes the BP algorithm applicable to more cases compared to the classic min-sum algorithm applied on a usual pairwise MRF [48].

Note that *reparameterization* (also known as *equivalent transformation*) of the MRF energy (*e.g.*, [181, 28]) is an important concept in MRF optimization. Two different settings of potentials (*e.g.*, θ_i, θ_{ij} in Eq. 11) leading to the same MRF energy (up to a constant difference) for any MRF configuration differ by a reparameterization. Reparameterization provides an alternative interpretation of belief propagation, which for example leads to a memory-efficient implementation of belief propagation [28]. Meanwhile, max-flow based algorithms also have been shown to relate to the principle of *reparameterization* [27]. Such a relationship (via reparameterization) sheds light on some connection between max-flow and message passing based algorithms.

4.2.2. Loopy Belief Propagation

The tree-structured constraint limits the use of the standard belief propagation algorithm presented above, whereas loopy MRFs are often required to model vision problems. Hence, researchers have investigated to extend the message passing concept for minimization of arbitrary graphs.

Loopy belief propagation (LBP), a natural step towards this direction, performs message passing iteratively in the graph (*e.g.*, [182, 48, 146, 147]) despite of the existence of loops. We refer the reader to [48, 146] for the details and discussion on the LBP algorithm. Regarding the message passing scheme in loopy graphs, there are two possible choices: *parallel* or *sequential*. In the parallel scheme, messages are computed for all the edges at the same time and then the messages are propagated for the

next round of message passing. Whereas in the sequential scheme, a node propagates the message to one of its neighbor node at each round and such a message will be used to compute the messages sent by that neighbor node. [183] showed empirically that the sequential scheme was significantly faster than the parallel one, while the performance of both methods was almost the same.

A number of works have been done to improve the efficiency of message passing by exploiting particular types of graphs and/or potential functions (*e.g.*, [147, 184, 185]). For example, based on the distance transform algorithm [186], [147] introduced a strategy for speeding up belief propagation for a subclass of pairwise potentials that only depend on the difference of the variables such as those defined in Eq. 13, which reduces the complexity of a message passing operation between two nodes from quadratic to linear in the number of possible labels per node. Techniques have also been proposed for accelerating the message passing in bipartite graphs and/or grid-like MRFs [147, 185], and in *robust truncated models* where a pairwise potential is equal to a constant for most of the possible state combinations of the two nodes [184]. Recently, [187] proposed a parallel message computation scheme, inspired from [147] but applicable to a wider subclass of MRFs than [147]. Together with a GPU implementation, such a scheme substantially reduces the running time in various MRF models for low-level vision problems.

Despite the fact that LBP performed well for a number of vision applications such as [48, 58], they cannot guarantee to converge to a fixed point, while their theoretical properties are not well understood. Last but not least, their solution is generally worse than more sophisticated generalizations of message passing algorithms (*e.g.*, [26, 28, 45]) that will be presented in section 4.3 [35].

4.2.3. Junction Tree Algorithm

Junction tree algorithm (JTA) is an exact inference method in arbitrary graphical models [138, 139, 12]. The key idea is to make systematic use of the Markov properties implied in graphical models to decompose a computation of the joint probability or energy into a set of local computations. Such an approach bears strong similarities with message passing in the standard belief propagation or dynamic programming. In

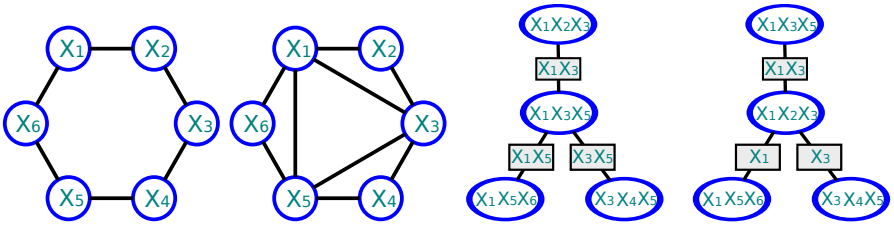


Figure 4: Example of Junction Tree. (a) Original undirected graphical model; (b) Triangulation of the graph in (a); (c) A junction tree for the graphs in (a) and (b); (d) A clique tree which is not junction tree. In (c-d), we use a square box to represent a separator being associated to an edge and denoting the intersection of the two cliques connected by the edge. A maximal spanning tree is a tree that connects all the nodes and has the maximal sum of the cardinals of the separators among all possible trees.

this sense, we regard JTA as an extension of the standard belief propagation.

An undirected graph has a *junction tree* if and only if it is triangulated (*i.e.*, there is no *chordless*[18] cycle in the graph). For any MRF, we can obtain a junction tree by first triangulating the original graph (*i.e.*, making the graph triangulated by adding additional edges) and then finding a maximal spanning tree for the maximal cliques contained in the triangulated graph (*e.g.*, Fig. 4). Based on the obtained junction tree, we can perform local message passing to do the exact inference, which is similar to standard belief propagation in factor trees. We refer the reader to [139, 12] for details.

The complexity of the inference in a junction tree for a discrete MRF is exponential with respect to its *width W*, which is defined as the maximum cardinal over all the maximal cliques minus 1. Hence, the complexity is dominated by the largest maximal cliques in the triangulated graph. However, the triangulation process may produce large maximal cliques, while finding of an optimal junction tree with the smallest width for an arbitrary undirected graph is an NP-hard problem. Furthermore, MRFs with dense initial connections could lead to maximal cliques of very high cardinal even if an optimal junction tree could be found [12]. Due to the computational complexity, the junction tree algorithm becomes impractical when the tree width is high, although it provides an exact inference approach. Thus it has been only used in some specific scenarios or some special kinds of graphs that have low tree widths (*e.g.*, cycles and

[18] A cycle is said to be *chordless* if there is no edge between two nodes that are not successors in the cycle.

outer-planar graphs whose widths are equal to 2). For example, JTA was employed in [179] to deal with simultaneous localization and mapping (SLAM) problem, and was also adopted in [92] to perform exactly inference in outer-planar graphs within the whole dual-decomposition framework. In order to reduce the complexity, *nested junction tree* technique was proposed in [188] to further factorize large cliques. Nevertheless, the gain of such a process depends directly on the initial graph structure and is still insufficient to make JTA widely applicable in practice.

4.3. Dual Methods

The MAP inference in pairwise MRFs (Eqs. 8 and 11), can be reformulated as the *integer linear programming (ILP)* [189] as follows:

$$\begin{aligned} \min_{\boldsymbol{\tau}} \quad & E(\boldsymbol{\theta},\boldsymbol{\tau}) = \langle \boldsymbol{\theta},\boldsymbol{\tau}\rangle = \sum_{i\in\mathcal{V}}\sum_{a\in\mathcal{X}_i}\theta_{i;a}\tau_{i;a} + \sum_{(i,j)\in\mathcal{E}}\sum_{(a,b)\in\mathcal{X}_i\times\mathcal{X}_j}\theta_{ij;ab}\tau_{ij;ab} \\ s.t. \quad & \boldsymbol{\tau}\in\boldsymbol{\tau}^{\mathcal{G}} = \left\{ \boldsymbol{\tau} \;\middle|\; \begin{array}{ll} \sum_{a\in\mathcal{X}_i}\tau_{i;a} = 1 & \forall\, i\in\mathcal{V} \\ \sum_{a\in\mathcal{X}_i}\tau_{ij;ab} = \tau_{j;b} & \forall\,\{i,j\}\in\mathcal{E}, b\in\mathcal{X}_j \\ \tau_{i;a}\in\{0,1\} & \forall\, i\in\mathcal{V}, a\in\mathcal{X}_i \\ \tau_{ij;ab}\in\{0,1\} & \forall\,\{i,j\}\in\mathcal{E}, (a,b)\in\mathcal{X}_i\times\mathcal{X}_j \end{array} \right\}. \end{aligned} \tag{18}$$

where $\theta_{i;a} = \theta_i(a)$, $\theta_{ij;ab} = \theta_{ij}(a,b)$, binary variables[19] $\tau_{i;a} = [x_u = a]$ and $\tau_{ij;ab} = [x_i = a, x_j = b]$, $\boldsymbol{\tau}$ denotes the concatenation of all these binary variables which can be defined as $((\tau_{i;a})_{i\in\mathcal{V},a\in\mathcal{X}_i}, (\tau_{ij;ab})_{\{i,j\}\in\mathcal{E},(a,b)\in\mathcal{X}_i\times\mathcal{X}_j})$, and $\boldsymbol{\tau}^{\mathcal{G}}$ denotes the domain of $\boldsymbol{\tau}$. We will use *MRF-MAP* to refer to this original MAP inference problem. Unfortunately, the above ILP problem is NP-hard in general[20]. Many approximation algorithms of MRF optimization have been developed based on solving some relaxation to such a problem.

Linear Programming (LP) relaxation has been widely adopted to address the MRF-MAP problem in Eq. 18, aiming to minimize $E(\boldsymbol{\theta},\boldsymbol{\tau})$ in a relaxed domain $\hat{\boldsymbol{\tau}}^{\mathcal{G}}$ (called *local marginal polytope*) which is obtained by simply replacing the integer constraints

[19] $[\cdot]$ is equal to one if the argument is true and zero otherwise.

[20] Note that, very recently, [190] experimentally demonstrated that for a subclass of small-size MRFs, advanced integer programming algorithms based on cutting-plane and branch-and-bound techniques can have global optimality property while being computational efficient.

in Eq. 18 by non-negative constraints (*i.e.*, $\tau_{i;a} \geq 0$ and $\tau_{ij;ab} \geq 0$). Such a relaxed problem will be referred to as *MRF-LP*. It is generally infeasible to directly apply generic LP algorithms such as *interior point methods* [191] to solve MRF-LP for MRF models in computer vision [192], due to the fact that the number of variables involved in $\boldsymbol{\tau}$ is usually huge. Instead, many methods have been designed based on solving some *dual* to MRF-LP, *i.e.*, maximizing the lower bound of $E(\boldsymbol{\theta}, \boldsymbol{\tau})$ provided by the dual. An important class of such methods are referred to as *tree-reweighted message passing (TRW)* techniques (*e.g.*, [26, 28]), which approach the solution to MRF-LP via a dual problem defined by a convex combination of trees. The optimal value of such a dual problem and that of MRF-LP coincide [26]. In [26], TRW was introduced to solve MRF-MAP by using edge-based and tree-based message passing schemes (called *TRW-E* and *TRW-T* respectively), which can be viewed as combinations of reparameterization and averaging operations on the MRF energy. However, the two schemes do not guarantee the convergence of the algorithms and the value of the lower bound may fall into a loop. Later, a sequential message passing scheme (known as *TRW-S*) was proposed in [28]. It updates messages in a sequential order instead of a parallel order used in TRW-E and TRW-T, which makes the lower bound will not decrease in TRW-S. Regarding the convergence, TRW-S will attain a point that satisfies a condition referred to as *weak tree agreement (WTA)* [193] and the lower bound will not change any more since then[21]. Regarding the optimality, TRW-S cannot guarantee the global maximum of the lower bound in general. Nevertheless, for the case of binary pairwise MRFs, a WTA fixed point corresponds to the global maximum of the lower bound, and thus the global minimum of MRF-LP [193]. Furthermore, if a binary pairwise MRF is submodular, a WTA fixed point always achieves the global optimum of the MRF-MAP problem. In [35], a set of experimental comparisons between ICM, LBP, α-expansion, $\alpha\beta$-swap and TRW-S were done based on MRFs with smoothness priors, showing that TRW-S and α-expansion perform much better than the others. For other representative methods solving a dual to MRF-LP, one can cite for example the *message passing* al-

[21] [28] observed in the experiments that TRW-S would finally converge to a fixed point but such a convergence required a lot of time after attaining WTA. Nevertheless, such a convergence may not be necessary in practice, since the lower bound will not change any more after attaining WTA.

gorithm based on *block coordinate descent* proposed in [194], the *min-sum diffusion* algorithm [195] and the *augmenting DAG* algorithm[22] [196], *etc*. Note that, since the LP-relaxation can be too loose to approach the solution of the MRF-MAP problem, the tightening of the LP-relaxation has also been investigated for achieving a better optimum of the MRF-MAP problem (*e.g*., [197, 198, 199, 30, 200, 201]).

Another important relaxation (*i.e*., Lagrangian relaxation) to MRF-MAP is related to *dual-decomposition* [202], which is a very important optimization methodology. *Dual-decomposition* was employed in [45, 156] for addressing the MRF-MAP problem (referred to as *MRF-DD*). The key idea is: instead of minimizing directly the energy of the original MRF-MAP problem which is too complex to solve directly, we decompose the original problem into a set of subproblems which are easy to solve. Based on a Lagrangian dual of the *MRF-MAP* problem, the sum of the minima of the subproblems provides a *lower bound* on the energy of the original MRF. This sum is maximized using *projected subgradient* method so that a solution to the original problem can be extracted from the Lagrangian solutions [156]. This leads to an MRF optimization framework with a high flexibility, generality and convergence property. First, the Lagrangian dual problem can be globally optimized due to the convexity of the dual function, which is a more desired property than WTA condition guaranteed by TRW-S. Second, different decompositions can be considered to deal with MRF-MAP, leading to different relaxations. In particular, when the master problem is decomposed into a set of trees, the obtained Lagrangian relaxation is equivalent to the LP relaxation of MRF-MAP. However, more sophisticated decompositions[23] can be considered to tighten the relaxation (*e.g*., decompositions based on outer-planar graphs [92] and k-fan graphs [91]). Third, there is no constraint on how the inference in slave problems is done and one can apply specific optimization algorithms to solve slave problems. A number of interesting applications have been proposed within such a framework, which include the graph matching method proposed in [203], the higher-order MRF inference method developed in [107], and the algorithm introduced in [204] for jointly inferring

[22] Both the *min-sum diffusion* algorithm and the *augmenting DAG* algorithm were reviewed in [155].

[23] A theoretical conclusion regarding the comparison the tightness between two different decompositions has been drawn in [156].

image segmentation and appearance histogram models. In addition, various techniques have been proposed to speed up the convergence of MRF-DD algorithms. For example, two approaches were introduced in [31]. One is to use a multi-resolution hierarchy of dual relaxations, and the other consists of a decimation strategy that gradually fixes the labels for a growing subset of nodes as well as their dual variables during the process. [205] proposed to construct a smooth approximation of the energy function of the master problem by smoothing the energies of the slave problems so as to achieve a significant acceleration of the MRF-DD algorithm. A distributed implementation of graph cuts was introduced in [206] to solve the slave problems in parallel.

Last, it is worth mentioning that an advantage of all dual methods is that we can tell how far the solution of MRF-MAP is from the global optimum, simply by measuring the gap between the lower bound obtained from solving the dual problem and the energy of the obtained MRF-MAP solution.

4.4. Inference in Higher-order MRFs

Recent development of higher-order MRF models for vision problems has been shown in section 3.2. In such a context, numerous works have been devoted in the past decade to search for efficient inference algorithms in higher-order models, towards expanding their use in vision problems that usually involve a large number of variables. One can cite for example [100, 101], where a simple inference scheme based on a conjugate gradient method was developed to solve their higher-order model for image restoration. Since then, besides a number of methods for solving specific types of higher-order models (*e.g.*, [102, 207, 118, 119, 122]), various techniques have also been proposed to deal with more general MRF models (*e.g.*, [208, 209, 107, 210, 211]). These inference methods are highly inspired from the ones for pairwise MRFs. Thus, similar to pairwise MRFs, there are also three main types of approaches for solving higher-order MRFs, *i.e.*, algorithms based on *order reduction* and *graph cuts*, higher-order extensions of *belief propagation*, and *dual methods*.

4.4.1. Order Reduction and Graph Cuts

Most of existing methods tackle inference in higher-order MRFs using a two-stage approach: first to reduce a higher-order model to a pairwise one with the same minimum, and then to apply standard methods such as graph cuts to solve the obtained pairwise model.

The idea of order reduction exists for long time. More than thirty years ago, a method (referred to as *variable substitution*) was proposed in [212] to perform order reduction for models of any order, by introducing auxiliary variables to substitute products of variables[24]. However, this approach leads to a large number of non-submodular components in the resulting pairwise model. This is due to the hard constraints involved in the substitution, which causes large difficulty in solving the obtained pairwise model. This may explain why its impact is rather limited in the literature [161, 213], since our final interest is solving higher-order models. In [213], QPBO was employed to solve the resulting pairwise model, nevertheless, only third-order potentials were tested in the experiments.

A better reduction method that generally produces fewer non-submodular components was proposed in [25], in order to construct s-t graph for a third-order binary MRF. This reduction method was studied from an algebraic viewpoint in [214] and led to some interesting conclusions towards extending this method to models of an arbitrary order. Based on these works, [210, 215] proposed a generalized technique that can reduce any higher-order binary MRF into a pairwise one, which can then be solved by QBPO. Furthermore, [210, 215] also extended such a technique to deal with multi-label MRFs by using fusion moves [172]. Very recently, aiming to obtain a pairwise model that is as easy as possible to solve (*i.e.*, has as few as possible non-submodular terms), [216] proposed to approach order reduction as an optimization problem, where different factors are allowed to choose different reduction methods in order to optimize an objective function defined using a special graph (referred to as *order reduction inference graph*). In the same line of research, [211] proposed to perform order reduction on

[24]Here, we consider binary higher-order MRFs and their energy functions can be represented in form of *pseudo-Boolean functions* [161].

a group of higher-order terms at the same time instead of on each term independently [210, 215], which has been demonstrated both theoretically and experimentally to lead to better performance compared to [210, 215].

Graph-cuts techniques have also been considered to cope either with specific vision problems or certain classes of higher-order models. For example, [102, 103] characterized a class of higher-order potentials (*i.e.*, $\mathcal{P}^n$ Potts model). It was also showed that the optimal expansion and swap moves for these higher-order potentials can be computed efficiently in polynomial time, which leads to an efficient graph-cuts-based algorithm for solving such models. Such a technique was further extended in [104, 105] to a wider class of higher-order models (*i.e.*, robust $\mathcal{P}^n$ model). In addition, graph-cuts-based approaches were also proposed in [122, 123, 119, 120, 217] to perform inference in their higher-order MRFs with global potentials that encode *co-occurrence statistics* and/or *label costs*. Despite the fact that such methods were designed for a limited range of problems that often cannot be solved by a general inference method, they better capture the characteristics of the problems and are able to solve the problems relatively efficiently.

4.4.2. Belief-propagation-based Methods

As mentioned in section 4.2, the factor graph representation of MRFs enables the extension of classic min-sum belief propagation algorithm to higher-order cases. Hence, loopy belief propagation in factor graphs provides a straightforward way to deal with inference in higher-order MRFs. Such an approach was adopted in [208] to solve their higher-order Fields-of-Experts model.

A practical problem for propagating messages in higher-order MRFs is that the complexity increases exponentially with respect to the highest order among all cliques. Various techniques have been proposed to accelerate the belief propagation in special families of higher-order potentials. For example, [218, 209] and [219] proposed efficient message passing algorithms for some families of potentials such as *linear constraint potentials* and *cardinality-based potentials*. Recently, the max-product message passing was accelerated in [220] by exploiting the fact that a clique potential often consists of a sum of potentials each involving only a sub-clique of variables, whose

expected computational time was further reduced in [221].

4.4.3. Dual Methods

The LP relaxation of the MRF-MAP problem for pairwise MRFs (see section 4.3) can be generalized to the cases of higher-order MRFs. Such a generalization was studied in [222, 200], where *min-sum diffusion* [195] was adopted to achieve a method for optimizing the energy of higher-order MRFs, which is referred to as *n-ary min-sum diffusion*[25]. Recently, such techniques were adopted in [223] to efficiently solve in a parallel/distributed fashion higher-order MRF models of triangulated planar structure.

The *dual-decomposition* framework [202, 154], which has been presented in section 4.3, can also be adopted to deal with higher-order MRFs. This was first demonstrated in [107], where inference algorithms were introduced for solving a wide class of higher-order potential referred to as *pattern-based potentials*[26]. Also based on the dual-decomposition framework, [115] proposed to solved their higher-order MRF model by decomposing the original problem into a series of subproblems each corresponding to a factor tree. In [224], such a framework was combined with order-reduction [210, 215] and QPBO techniques [150] to solve higher-order graph-matching problems.

Exploitation of the Sparsity of Potentials

Last, it is worth mentioning that the sparsity of potentials has been exploited, either explicitly or implicitly, in many of the above higher-order inference methods. For example, [225] proposed a compact representation for "sparse" higher-order potentials (except a very small subset, the labelings are almost impossible and have the same high energy), via which a higher-order model can be converted into a pairwise one by introducing only a small number of auxiliary variables and then pairwise MRF inference methods such as graph cuts can be employed to solve the problem. In the same line of research, [226] studied and characterized some classes of higher-order potentials (*e.g.*, $\mathcal{P}^n$ Potts model [103]) that can be represented compactly as upper or lower envelope

[25] The method was originally called *n-ary max-sum diffusion* in [222, 200] due to the fact that a maximization of objective function was considered.

[26] For example, $\mathcal{P}^n$ Potts model [103] is a sub-class of *pattern-based potentials*.

of linear functions. Furthermore, it was demonstrated in [226] that these higher-order models can be converted into pairwise models with the addition of a small number of auxiliary variables. [227] proposed to optimize the energy of "sparse" higher-order models by transforming the original problem into a relatively small instance of submodular vertex-cover, which can then be optimized by standard algorithms such as belief propagation and QPBO. This approach has been shown to achieve much better efficiency than applying those standard algorithms to address the original problem directly. Very recently, [228] took a further step along this line of research by exploring the intrinsic dimensions of higher-order cliques, and proposed a powerful MRF-based modeling/inference framework (called *NC-MRF*) which significantly broadens the applicability of higher-order MRFs in visual perception.

5. MRF Learning Methods

On top of inference, another task of great importance is MRF learning/training, which aims to select the optimal model from its feasible set based on the training data. In this case, the input is a set of K training samples $\{\mathbf{d}^k, \mathbf{x}^k\}_{k=1}^K$, where $\mathbf{d}^k$ and $\mathbf{x}^k$ represent the observed data and the ground truth MRF configuration of the k-th sample, respectively. Moreover, it is assumed that the unary potentials θ_i^k and the pairwise potentials θ_{ij}^k of the k-th MRF training instance can be expressed linearly in terms of feature vectors extracted from the observed data $\mathbf{d}^k$, that is, it holds $\theta_i^k(x_i) = \mathbf{w}^T g_i(x_i, \mathbf{d}^k)$, $\theta_{ij}^k(x_i, x_j) = \mathbf{w}^T g_{ij}(x_i, x_j, \mathbf{d}^k)$, where $g_i(\cdot,\cdot)$ and $g_{ij}(\cdot,\cdot)$ represent some known vector-valued feature functions (which are chosen based on the computer vision application at hand) and $\mathbf{w}$ is an unknown vector of parameters. The goal of MRF learning boils down to estimating this vector $\mathbf{w}$ using as input the above training data.

Both generative (*e.g.*, maximum-likelihood) and discriminative (*e.g.*, max-margin) MRF learning approaches have been applied for this purpose. In the former case, one seeks to maximize (possibly along with an L2-norm regularization term) the product of posterior probabilities of the ground truth MRF labelings $\prod_k P(\mathbf{x}^k; \mathbf{w})$, where $P(\mathbf{x}; \mathbf{w}) \propto \exp(E(\mathbf{x}; \mathbf{w}))$ denotes the probability distribution induced by an MRF model with energy $E(\mathbf{x}; \mathbf{w})$. This leads to a convex differentiable objective function that can

be optimized using gradient ascent. However, computing the gradient of this function involves taking expectations of the feature functions, $g_i(\cdot)$ and $g_{ij}(\cdot)$, with respect to the MRF distribution $P(\mathbf{x};\mathbf{w})$. One therefore needs to perform probabilistic MRF inference, which is nevertheless intractable in general. As a result, approximate inference techniques (*e.g.*, loopy belief propagation) are often used for approximating the MRF marginals required for the estimation of the gradient. This is the case, for instance, in [5], where the authors demonstrate how to train a CRF model for stereo matching, as well as in [3], or in [2], where a comparison with other CRF training methods such as pseudo-likelihood and MCMC-based contrastive divergence is also included.

In the case of max-margin learning [229, 230], on the other hand, one seeks to adjust the vector $\mathbf{w}$ such that the energy $E(\mathbf{x}^k;\mathbf{w})$ of the desired ground truth solution $\mathbf{x}^k$ is smaller by $\Delta(\mathbf{x},\mathbf{x}^k)$ than the energy $E(\mathbf{x};\mathbf{w})$ of any other solution $\mathbf{x}$, that is,

$$E(\mathbf{x}^k;\mathbf{w}) \leq E(\mathbf{x};\mathbf{w}) - \Delta(\mathbf{x},\mathbf{x}^k) + \xi_k \quad . \tag{19}$$

In the above set of linear inequality constraints with respect to $\mathbf{w}$, $\Delta(\mathbf{x},\mathbf{x}')$ represents a user-specified distance function that measures the dissimilarity between any two solutions $\mathbf{x}$ and $\mathbf{x}'$ (obviously it should hold $\Delta(\mathbf{x},\mathbf{x}) = 0$), while ξ_k is a non-negative slack variable that has been introduced for ensuring that a feasible solution $\mathbf{w}$ does exist. The distance function $\Delta(\mathbf{x},\mathbf{x}')$ modulates the margin according to how "far" an MRF labeling differs from the ground truth labeling. In practice, its choice is largely constrained by the tractability of the whole learning algorithm. The Hamming distance is often used in the literature [231, 232], due to the fact that it can be decomposed into a sum of unary terms and integrated easily in the MRF energy without increasing the order of the MRF model. However, visual perception often prefers more sophisticated task-specific distances that can better characterize the physical meaning of the labeling. For example, [233, 234] have investigated the incorporation of various higher-order distance functions in MRF learning for the image segmentation task.

Ideally, $\mathbf{w}$ should be set such that each $\xi_k \geq 0$ can take a value as small as possible (so that the amount of violation of the above constraints is minimal). As a result, during

the MRF learning, the following constrained optimization problem is solved:

$$\min_{\mathbf{w},\{\xi_k\}} \ \mu \cdot R(\mathbf{w}) + \sum_{k=1}^{K} \xi_k, \quad \text{s.t. constraints (19)} \ . \tag{20}$$

In the above problem, μ is a user-specified hyperparameter and $R(\mathbf{w})$ represents a regularization term whose role is to prevent overfitting during the learning process (*e.g.*, it can be set equal to $\|\mathbf{w}\|^2$ or to a sparsity inducing norm such as $\|\mathbf{w}\|_1$). The slack variable ξ_k can also be expressed as the following hinge-loss term:

$$\text{Loss}(\mathbf{x}^k; \mathbf{w}) = E(\mathbf{x}^k; \mathbf{w}) - \min_{\mathbf{x}} \Big(E(\mathbf{x}; \mathbf{w}) - \Delta(\mathbf{x}, \mathbf{x}^k) \Big) \ . \tag{21}$$

This leads to the following equivalent unconstrained formulation:

$$\min_{\mathbf{w}} \ \mu \cdot R(\mathbf{w}) + \sum_{k=1}^{K} \text{Loss}(\mathbf{x}^k; \mathbf{w}) \ . \tag{22}$$

One class of methods [235, 236] aim to solve the constrained optimization problem (20) by the use of a cutting-plane approach when $R(\mathbf{w}) = \|\mathbf{w}\|^2$. In this case, the above problem is equivalent to a convex quadratic program (QP) but with an exponential number of linear inequality constraints. Given that only a small fraction of them will be active at an optimal solution, cutting plane methods proceed by solving a small QP with a growing number of constraints at each iteration (where this number is polynomially upper-bounded). One drawback of such an approach relates to the fact that computing a violated constraint requires solving at each iteration a MAP inference problem that is NP-hard in general. For the special case of submodular MRFs, [237] shows how to express the above constraints (19) in a compact form, which allows for a more efficient MRF learning to take place in this case.

Another class of methods tackle instead the unconstrained formulation (22). This is, *e.g.*, the case for the recently proposed framework by [238], which addresses the above mentioned drawbacks of the cutting plane method by relying on the dual decomposition approach for MRF-MAP inference discussed previously in section 4.3. By using such an approach, this framework reduces the task of training an arbitrar-

ily complex MRF to that of training in parallel a series of simpler slave MRFs that are much easier to handle within a max-margin framework. The concurrent training of the slave MRFs takes place through a very efficient stochastic subgradient learning scheme. Moreover, such a framework can efficiently handle not only pairwise but also high-order MRFs, as well as any convex regularizer $R(\mathbf{w})$.

There have also been developed learning methods [239, 240, 241] that aim to deal with the training of MRFs that contain *latent* variables, *i.e.*, variables that remain unknown during both training and testing. Such MRF models are often encountered in vision applications due to the fact that in many cases full annotation is difficult or at least very time consuming to be provided (especially for large scale datasets). As a result, one often has to deal with datasets that are only partially annotated (weakly supervised learning).

Last but not least, there have also been proposed learning algorithms that are appropriate for handling the discriminative training of continuous MRF models [242].

6. Conclusion

In order to conclude this survey, let us first recall that developing MRF-based methods for vision problems and efficient inference algorithms has been a dominant research direction in computer vision during the past decade. The main stream referred to pairwise formulations, whereas more and more focus has been recently transferred to higher-order MRFs in order to achieve superior solutions for a wider set of vision problems. Moreover, machine learning techniques have been combined more and more with MRFs towards image/scene understanding as well as parameter learning and structure learning of MRF models. All these suggest that MRFs will keep being a major research topic and offer more promise than ever before.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments. Part of the work was done while C. Wang was with the Vision Lab at

University of California, Los Angeles, USA. N. Paragios' work was partially supported from the European Research Council Starting Grant DIOCLES (ERC-STG-259112).

Appendix A. Submodularity of MRFs

There are various definitions of *submodular* energy functions of pairwise discrete MRFs in the literature that are equivalent. We consider here the one presented in [140]. Let us assume the configuration space $\mathcal{X}_i$ for a node $i \in \mathcal{V}$ to be a completely ordered set, the energy function of a pairwise discrete MRF is *submodular* if each pairwise potential term θ_{ij} $(\forall \{i, j\} \in \mathcal{E})$ satisfies: $\forall\, x_i^1, x_i^2 \in \mathcal{X}_i \; s.t. \; x_i^1 \leq x_i^2$, and $\forall\, x_j^1, x_j^2 \in \mathcal{X}_j \; s.t. \; x_j^1 \leq x_j^2$,

$$\theta_{ij}(x_i^1, x_j^1) + \theta_{ij}(x_i^2, x_j^2) \leq \theta_{ij}(x_i^1, x_j^2) + \theta_{ij}(x_i^2, x_j^1) \ . \tag{A.1}$$

For binary cases where $\mathcal{X}_i = \{0, 1\}$ $(\forall\, i \in \mathcal{V})$, the condition is reduced to that each pairwise potential θ_{ij} $(\forall \{i, j\} \in \mathcal{E})$ satisfy:

$$\theta_{ij}(0, 0) + \theta_{ij}(1, 1) \leq \theta_{ij}(0, 1) + \theta_{ij}(1, 0) \ . \tag{A.2}$$

One can refer to [25] for generalizing the submodularity to higher-order MRFs.

Appendix B. Min-sum Belief Propagation in Factor Tree

Algorithm 1 Min-sum Belief Propagation in Factor Tree

Require: Factor tree $\mathcal{T} = (\mathcal{V} \cup \mathcal{F}, \mathcal{E})$ with usual node set $\mathcal{V}$, factor node set $\mathcal{F}$ and edge set $\mathcal{E}$
Require: Factor potentials $(\theta_f(\cdot))_{f \in \mathcal{F}}$
Ensure: The optimal configuration $\mathbf{x}^{\text{opt}} = \text{argmin}_{\mathbf{x}} \sum_{f \in \mathcal{F}} \theta_f(x_f)$
  Choose a node $\hat{r} \in \mathcal{V}$ as the root of the tree
  Construct Π *s.t.* $\Pi(i)$ denotes the parent of node $i \in \mathcal{V} \cup \mathcal{F}$
  Construct $\boldsymbol{C}$ *s.t.* $\boldsymbol{C}(i)$ denotes the set of children of node $i \in \mathcal{V} \cup \mathcal{F}$
  $\mathcal{P}_{\text{send}} \leftarrow \text{NodeOrdering}(\mathcal{T}, \hat{r})$ {see Algorithm 2}
  for $k = 1 \rightarrow \text{length}(\mathcal{P}_{\text{send}}) - 1$ **do**
    $i \leftarrow \mathcal{P}_{\text{send}}(k)$
    parent node $p \leftarrow \Pi(i)$
    child node set $C \leftarrow \boldsymbol{C}(i)$
    if $i \in \mathcal{V}$ **then**
      if $|C| > 0$ **then**
        $m_{i \to p}(x_i) \leftarrow \sum_{j \in C} m_{j \to i}(x_i)$
      else
        $m_{i \to p}(x_i) \leftarrow 0$
      end if
    else
      if $|C| > 0$ **then**
        $m_{i \to p}(x_p) \leftarrow \min_{x_C}(\phi(x_i) + \sum_{j \in C} m_{j \to i}(x_j))$
        $s_i(x_p) \leftarrow \text{argmin}_{x_C}(\phi(x_i) + \sum_{j \in C} m_{j \to i}(x_j))$
      else
        $m_{i \to p}(x_p) \leftarrow \phi(x_p)$ {p is the unique variable contained in factor i in this case.}
      end if
    end if
  end for
  $x_{\hat{r}}^{\text{opt}} \leftarrow \text{argmin}_{x_{\hat{r}}} \sum_{j \in C(\hat{r})} m_{j \to \hat{r}}(x_{\hat{r}})$
  for $k = \text{length}(\mathcal{P}_{\text{send}}) - 1 \rightarrow 1$ **do**
    $i \leftarrow \mathcal{P}_{\text{send}}(k)$
    if $i \in \mathcal{F}$ **then**
      parent node $p \leftarrow \Pi(i)$
      child node set $C \leftarrow \boldsymbol{C}(i)$
      $x_C^{\text{opt}} \leftarrow s_i(x_p)$
    end if
  end for
  return $\mathbf{x}^{\text{opt}}$

Algorithm 2 Ordering of the Nodes for Sending Messages In a Tree

Require: Tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with node set $\mathcal{V}$ and edge set $\mathcal{E}$

Require: Root node $\hat{r} \in \mathcal{V}$

Ensure: $\mathcal{P}_{\text{send}}$ = NodeOrdering($\mathcal{T}, \hat{r}$), where $\mathcal{P}_{\text{send}}$ is a list denoting the ordering of the nodes in tree $\mathcal{T}$ for sending messages

$\mathcal{P}_{\text{send}} \leftarrow (\hat{r})$

if $|\mathcal{V}| > 1$ **then**

Get the set $\mathcal{C}$ of child nodes: $\mathcal{C} \leftarrow \{i | i \in \mathcal{V}, \{i, \hat{r}\} \in \mathcal{E}\}$

for all $c \in \mathcal{C}$ **do**

Get child tree $\mathcal{T}_c$ with root c

$\mathcal{P}_{\text{send}} \leftarrow$ (NodeOrdering($\mathcal{T}, \hat{r}$), $\mathcal{P}_{\text{send}}$) {$\mathcal{P}_{\text{send}}$ is ordered from left to right}

end for

end if

return $\mathcal{P}_{\text{send}}$

References

[1] R. Szeliski, Computer Vision: Algorithms and Applications, Springer-Verlag New York Inc., 2010.

[2] S. Kumar, J. August, M. Hebert, Exploiting Inference for Approximate Parameter Learning in Discriminative Fields: An Empirical Study, in: International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2005.

[3] S. Kumar, M. Hebert, Discriminative Random Fields, International Journal of Computer Vision 68 (2) (2006) 179–201.

[4] S. Roth, M. J. Black, On the Spatial Statistics of Optical Flow, International Journal of Computer Vision (IJCV) 74 (1) (2007) 33–50.

[5] D. Scharstein, C. Pal, Learning Conditional Random Fields for Stereo, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[6] P. Kohli, P. H. S. Torr, Measuring Uncertainty in Graph Cut Solutions, Computer Vision and Image Understanding (CVIU) 112 (1) (2008) 30–38.

[7] D. Tarlow, R. P. Adams, Revisiting Uncertainty in Graph Cut Solutions, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[8] A. N. Tikhonov, V. Y. Arsenin, Solutions of Ill-posed Problems, Winston Washington, DC:, 1977.

[9] H. W. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer Academic Publishers, Dordrecht, 1996.

[10] S. L. Lauritzen, Graphical Models, Oxford University Press, 1996.

[11] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, 2006.

[12] M. I. Jordan, An Introduction to Probabilistic Graphical Models, In preparation, 2007.

[13] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.

[14] J. M. Hammersley, P. Clifford, Markov Fields on Finite Graphs and Lattices, unpublished.

[15] J. Besag, Spatial Interaction and the Statistical Analysis of Lattice Systems, Journal of the Royal Statistical Society. Series B (Methodological) 36 (2) (1974) 192–236.

[16] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to Algorithms (Third Edition), MIT Press, 2009.

[17] S. Z. Li, Markov Random Field Modeling in Image Analysis (Third Edition), Springer, 2009.

[18] A. Blake, P. Kohli, C. Rother (Eds.), Markov Random Fields for Vision and Image Processing, MIT Press, 2011.

[19] M. Isard, PAMPAS: Real-valued Graphical Models for Computer Vision, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2003.

[20] L. Sigal, M. Isard, B. H. Sigelman, M. J. Black, Attractive People: Assembling Loose-Limbed Models using Non-parametric Belief Propagation, in: Advances in Neural Information Processing Systems (NIPS), 2003.

[21] E. B. Sudderth, A. T. Ihler, M. Isard, W. T. Freeman, A. S. Willsky, Nonparametric Belief Propagation, Communications of the ACM 53 (10) (2010) 95–103.

[22] H. Rue, L. Held, Gaussian Markov Random Fields: Theory and Applications, Chapman & HALL/CRC, 2005.

[23] Y. Boykov, O. Veksler, R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 23 (11) (2001) 1222–1239.

[24] H. Ishikawa, Exact Optimization for Markov Random Fields with Convex Priors, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 25 (10) (2003) 1333–1336.

[25] V. Kolmogorov, R. Zabih, What Energy Functions can be Minimized via Graph Cuts?, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 26 (2) (2004) 147–159.

[26] M. J. Wainwright, T. S. Jaakkola, A. S. Willsky, MAP Estimation via Agreement on Trees: Message-passing and Linear Programming, IEEE Transactions on Information Theory 51 (11) (2005) 3697–3717.

[27] P. Kohli, P. H. S. Torr, Dynamic Graph Cuts for Efficient Inference in Markov Random Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 29 (12) (2007) 2079–2088.

[28] V. Kolmogorov, Convergent Tree-reweighted Message Passing for Energy Minimization, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 28 (10) (2006) 1568–1583.

[29] N. Komodakis, G. Tziritas, N. Paragios, Performance vs Computational Efficiency for Optimizing Single and Dynamic MRFs: Setting the State of the Art with Primal-Dual Strategies, Computer Vision and Image Understanding (CVIU) 112 (1) (2008) 14–29.

[30] M. Pawan Kumar, V. Kolmogorov, P. H. S. Torr, An Analysis of Convex Relaxations for MAP Estimation of Discrete MRFs, Journal of Machine Learning Research 10 (2009) 71–106.

[31] N. Komodakis, Towards More Efficient and Effective LP-Based Algorithms for MRF Optimization, in: European Conference on Computer Vision (ECCV), 2010.

[32] V. Kolmogorov, R. Zabih, Multi-camera Scene Reconstruction via Graph Cuts, in: European Conference on Computer Vision (ECCV), 2002.

[33] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, N. Paragios, Dense Image Registration through MRFs and Efficient Linear Programming, Medical Image Analysis 12 (6) (2008) 731–741.

[34] P. Kohli, J. Rihan, M. Bray, P. H. S. Torr, Simultaneous Segmentation and Pose Estimation of Humans Using Dynamic Graph Cuts, International Journal of Computer Vision (IJCV) 79 (3) (2008) 285–298.

[35] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-based Priors, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 30 (6) (2008) 1068–1080.

[36] Y. Boykov, G. Funka-Lea, Graph Cuts and Efficient N-D Image Segmentation, International Journal of Computer Vision (IJCV) 70 (2) (2006) 109–131.

[37] B. J. Frey, Graphical Models for Machine Learning and Digital Communication, MIT Press, 1998.

[38] F. R. Kschischang, B. J. Frey, H.-A. Loeliger, Factor Graphs and the Sum-Product Algorithm, IEEE Transactions on Information Theory 47 (2) (2001) 498–519.

[39] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, 1988.

[40] J. S. Yedidia, W. T. Freeman, Y. Weiss, Understanding Belief Propagation and its Generalizations, in: Exploring Artificial Intelligence in the New Millennium, Morgan Kaufmann, 2003, pp. 239–269.

[41] S. Geman, D. Geman, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 6 (6) (1984) 721–741.

[42] C. Rother, V. Kolmogorov, A. Blake, GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts, ACM Transactions on Graphics (TOG) 23 (3) (2004) 309–314.

[43] P. F. Felzenszwalb, D. P. Huttenlocher, Pictorial Structures for Object Recognition, International Journal of Computer Vision (IJCV) 61 (1) (2005) 55–79.

[44] B. Glocker, A. Sotiras, N. Komodakis, N. Paragios, Deformable Medical Image Registration: Setting the State of the Art with Discrete Methods, Annual Review of Biomedical Engineering 13 (1) (2011) 219–244.

[45] N. Komodakis, N. Paragios, G. Tziritas, MRF Optimization via Dual Decomposition: Message-Passing Revisited, in: IEEE International Conference on Computer Vision (ICCV), 2007.

[46] D. M. Greig, B. T. Porteous, A. H. Seheult, Exact Maximum A Posteriori Estimation for Binary Images, Journal of the Royal Statistical Society (Series B) 51 (2) (1989) 271–279.

[47] A. Chambolle, Total Variation Minimization and a Class of Binary MRF Models, in: International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2005.

[48] W. T. Freeman, E. C. Pasztor, O. T. Carmichael, Learning Low-level Vision, International Journal of Computer Vision (IJCV) 40 (1) (2000) 25–47.

[49] W. T. Freeman, T. R. Jones, E. C. Pasztor, Example-Based Super-Resolution, IEEE Computer Graphics and Applications 22 (2) (2002) 56–65.

[50] D. Rajan, S. Chaudhuri, An MRF-Based Approach to Generation of Super-Resolution Images from Blurred Observations, Journal of Mathematical Imaging and Vision 16 (1) (2002) 5–15.

[51] S. Roy, I. J. Cox, A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem, in: IEEE International Conference on Computer Vision (ICCV), 1998.

[52] G. Vogiatzis, C. H. Esteban, P. H. S. Torr, R. Cipolla, Multiview Stereo via Volumetric Graph-Cuts and Occlusion Robust Photo-Consistency, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 29 (12) (2007) 2241–2246.

[53] F. Heitz, P. Bouthemy, Multimodal Estimation of Discontinuous Optical Flow using Markov Random Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 15 (12) (1993) 1217–1232.

[54] S. Roy, V. Govindu, MRF Solutions for Probabilistic Optical flow Formulations, in: International Conference on Pattern Recognition (ICPR), 2000.

[55] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, N. Navab, Optical Flow Estimation with Uncertainties through Dynamic MRFs, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[56] C. Liu, J. Yuen, A. Torralba, SIFT Flow: Dense Correspondence across Scenes and Its Applications, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 33 (5) (2011) 978–994.

[57] B. Glocker, N. Komodakis, N. Navab, G. Tziritas, N. Paragios, Dense Registration with Deformation Priors, in: International Conference on Information Processing in Medical Imaging (IPMI), 2009.

[58] J. Sun, N.-N. Zheng, H.-Y. Shum, Stereo Matching using Belief Propagation, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 25 (7) (2003) 787–800.

[59] A. Shekhovtsov, I. Kovtun, V. Hlavac, Efficient MRF Deformation Model for Non-rigid Image Matching, Computer Vision and Image Understanding (CVIU) 112 (1) (2008) 91–99.

[60] Y. Boykov, V. Kolmogorov, Computing Geodesics and Minimal Surfaces via Graph Cuts, in: IEEE International Conference on Computer Vision (ICCV), 2003.

[61] D. Singaraju, L. Grady, R. Vidal, P-brush: Continuous Valued MRFs with Normed Pairwise Distributions for Image Segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[62] A. P. Moore, S. J. D. Prince, J. Warrell, "Lattice Cut" - Constructing Superpixels using Layer Constraints, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[63] O. Veksler, Y. Boykov, P. Mehrani, Superpixels and Supervoxels in an Energy Optimization Framework, in: European Conference on Computer Vision (ECCV), 2010.

[64] Y. Zhang, R. Hartley, J. Mashford, S. Burn, Superpixels via Pseudo-Boolean Optimization, in: IEEE International Conference on Computer Vision (ICCV), 2011.

[65] E. Ising, Beitrag zur theorie des ferromagnetismus, Zeitschrift fur Physik 31 (1) (1925) 253–258.

[66] R. B. Potts, Some Generalized Order-disorder Transitions, Proceedings of the Cambridge Philosophical Society 48 (1952) 106–109.

[67] Y. Boykov, O. Veksler, R. Zabih, Markov Random Fields with Efficient Approximations, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1998.

[68] D. Terzopoulos, Regularization of Inverse Visual Problems Involving Discontinuities, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 8 (4) (1986) 413–424.

[69] D. Lee, T. Pavlidis, One-dimensional Regularization with Discontinuities, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 10 (6) (1988) 822–829.

[70] O. Veksler, Star Shape Prior for Graph-Cut Image Segmentation, in: European Conference on Computer Vision (ECCV), 2008.

[71] P. Das, O. Veksler, V. Zavadsky, Y. Boykov, Semiautomatic Segmentation with Compact Shape Prior, Image and Vision Computing (IVC) 27 (1-2) (2009) 206–219.

[72] F. R. Schmidt, Y. Boykov, Hausdorff Distance Constraint for Multi-surface Segmentation, in: European Conference on Computer Vision (ECCV), 2012.

[73] X. Liu, O. Veksler, J. Samarabandu, Order-preserving Moves for Graph-cut-based Optimization, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 32 (7) (2010) 1182–1196.

[74] J. Bai, Q. Song, O. Veksler, X. Wu, Fast Dynamic Programming for Labeling Problems with Ordering Constraints, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[75] B. Fulkerson, A. Vedaldi, S. Soatto, Class Segmentation and Object Localization with Superpixel Neighborhoods, in: IEEE International Conference on Computer Vision (ICCV), 2009.

[76] A. Levinshtein, C. Sminchisescu, S. Dickinson, Optimal Contour Closure by Superpixel Grouping, in: European Conference on Computer Vision (ECCV), 2010.

[77] E. Kalogerakis, A. Hertzmann, K. Singh, Learning 3D Mesh Segmentation and Labeling, ACM Transactions on Graphics (TOG) 29 (4) (2010) 102:1–102:12.

[78] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, N. Paragios, Intrinsic Dense 3D Surface Tracking, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[79] L. Sigal, A. O. Balan, M. J. Black, HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion, International Journal of Computer Vision (IJCV) 87 (1-2) (2010) 4–27.

[80] M. Fischler, R. Elschlager, The Representation and Matching of Pictorial Structures, IEEE Transactions on Computers 22 (1) (1973) 67–92.

[81] R. Bellman, Dynamic Programming, Princeton University Press, 1957.

[82] L. Sigal, M. J. Black, Measure Locally, Reason Globally: Occlusion-sensitive Articulated Pose Estimation, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006.

[83] M. Eichner, V. Ferrari, Better Appearance Models for Pictorial Structures, in: British Machine Vision Conference (BMVC), 2009.

[84] M. Andriluka, S. Roth, B. Schiele, Pictorial Structures Revisited: People Detection and Articulated Pose Estimation, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[85] L. Pishchulin, M. Andriluka, P. Gehler, B. Schiele, Poselet Conditioned Pictorial Structures, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[86] E. B. Sudderth, M. I. Mandel, W. T. Freeman, A. S. Willsky, Visual Hand Tracking Using Nonparametric Belief Propagation, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops), 2004.

[87] E. B. Sudderth, M. I. Mandel, W. T. Freeman, A. S. Willsky, Distributed Occlusion Reasoning for Tracking with Nonparametric Belief Propagation, in: Advances in Neural Information Processing Systems (NIPS), 2004.

[88] M. Pawan Kumar, P. H. S. Torr, A. Zisserman, Learning Layered Pictorial Structures from Video, in: The Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP), 2004.

[89] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan, Object Detection with Discriminatively Trained Part-based Models, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 32 (9) (2010) 1627–1645.

[90] D. Crandall, P. Felzenszwalb, D. Huttenlocher, Spatial Priors for Part-Based Recognition Using Statistical Models, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[91] J. H. Kappes, S. Schmidt, C. Schnorr, MRF Inference by k-Fan Decomposition and Tight Lagrangian Relaxation, in: European Conference on Computer Vision (ECCV), 2010.

[92] D. Batra, A. C. Gallagher, D. Parikh, T. Chen, Beyond Trees: MRF Inference via Outer-Planar Decomposition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[93] L. Bourdev, J. Malik, Poselets: Body Part Detectors Trained using 3D Human Pose Annotations, in: IEEE International Conference on Computer Vision (ICCV), 2009.

[94] C. Wang, M. de La Gorce, N. Paragios, Segmentation, Ordering and Multi-object Tracking Using Graphical Models, in: IEEE International Conference on Computer Vision (ICCV), 2009.

[95] T. Heimann, H.-P. Meinzer, Statistical Shape Models for 3D Medical Image Segmentation: A Review, Medical Image Analysis 13 (4) (2009) 543–563.

[96] D. Seghers, D. Loeckx, F. Maes, D. Vandermeulen, P. Suetens, Minimal Shape and Intensity Cost Path Segmentation, IEEE Transactions on Medical Imaging (TMI) 26 (8) (2007) 1115–1129.

[97] A. Besbes, N. Komodakis, G. Langs, N. Paragios, Shape Priors and Discrete MRFs for Knowledge-based Segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[98] T. H. Heibel, B. Glocker, M. Groher, N. Paragios, N. Komodakis, N. Navab, Discrete Tracking of Parametrized Curves, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[99] B. Xiang, C. Wang, J.-F. Deux, A. Rahmouni, N. Paragios, Tagged Cardiac MR Image Segmentation Using Boundary & Regional-Support and Graph-based Deformable Priors, in: IEEE International Symposium on Biomedical Imaging (ISBI), 2011.

[100] S. Roth, M. J. Black, Fields of Experts: A Framework for Learning Image Priors, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[101] S. Roth, M. J. Black, Fields of Experts, International Journal of Computer Vision (IJCV) 82 (2) (2009) 205–229.

[102] P. Kohli, M. Pawan Kumar, P. H. S. Torr, P3 & Beyond: Solving Energies with Higher Order Cliques, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[103] P. Kohli, M. Pawan Kumar, P. H. S. Torr, P3 & Beyond: Move Making Algorithms for Solving Higher Order Functions, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 31 (9) (2009) 1645–1656.

[104] P. Kohli, L. Ladický, P. H. S. Torr, Robust Higher Order Potentials for Enforcing Label Consistency, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[105] P. Kohli, L. Ladický, P. H. S. Torr, Robust Higher Order Potentials for Enforcing Label Consistency, International Journal of Computer Vision (IJCV) 82 (3) (2009) 302–324.

[106] O. J. Woodford, P. H. S. Torr, I. D. Reid, A. W. Fitzgibbon, Global Stereo Reconstruction under Second-Order Smoothness Priors, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 31 (12) (2009) 2115–2128.

[107] N. Komodakis, N. Paragios, Beyond Pairwise Energies: Efficient Optimization for Higher-order MRFs, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[108] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active Contour Models, International Journal of Computer Vision (IJCV) 1 (4) (1988) 321–331.

[109] A. A. Amini, T. E. Weymouth, R. C. Jain, Using Dynamic Programming for Solving Variational Problems in Vision, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 12 (9) (1990) 855–867.

[110] D. Kwon, K. J. Lee, I. D. Yun, S. U. Lee, Nonrigid Image Registration Using Dynamic Higher-Order MRF Model, in: European Conference on Computer Vision (ECCV), 2008.

[111] B. Glocker, T. H. Heibel, N. Navab, P. Kohli, C. Rother, TriangleFlow: Optical Flow with Triangulation-Based Higher-Order Likelihoods, in: European Conference on Computer Vision (ECCV), 2010.

[112] A. Shekhovtsov, P. Kohli, C. Rother, Curvature Prior for MRF-Based Segmentation and Shape Inpainting, in: DAGM/OAGM Symposium, 2012.

[113] V. Lempitsky, P. Kohli, C. Rother, T. Sharp, Image Segmentation with a Bounding Box Prior, in: IEEE International Conference on Computer Vision (ICCV), 2009.

[114] A. Panagopoulos, C. Wang, D. Samaras, N. Paragios, Simultaneous Cast Shadows, Illumination and Geometry Inference Using Hypergraphs, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 35 (2) (2013) 437–449.

[115] C. Wang, O. Teboul, F. Michel, S. Essafi, N. Paragios, 3D Knowledge-Based Segmentation Using Pose-Invariant Higher-Order Graphs, in: International Conference, Medical Image Computing and Computer Assisted Intervention (MICCAI), 2010.

[116] C. Wang, Y. Zeng, L. Simon, I. Kakadiaris, D. Samaras, N. Paragios, Viewpoint Invariant 3D Landmark Model Inference from Monocular 2D Images Using Higher-Order Priors, in: IEEE International Conference on Computer Vision (ICCV), 2011.

[117] S. Vicente, V. Kolmogorov, C. Rother, Graph Cut Based Image Segmentation with Connectivity Priors, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[118] S. Nowozin, C. H. Lampert, Global Connectivity Potentials for Random Field Models, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[119] A. Delong, A. Osokin, H. N. Isack, Y. Boykov, Fast Approximate Energy Minimization with Label Costs, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[120] A. Delong, A. Osokin, H. N. Isack, Y. Boykov, Fast Approximate Energy Minimization with Label Costs, International Journal of Computer Vision (IJCV) 96 (1) (2012) 1–27.

[121] S. C. Zhu, A. Yuille, Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 18 (9) (1996) 884–900.

[122] L. Ladický, C. Russell, P. Kohli, P. H. S. Torr, Graph Cut based Inference with Co-occurrence Statistics, in: European Conference on Computer Vision (ECCV), 2010.

[123] L. Ladický, C. Russell, P. Kohli, P. H. S. Torr, Inference Methods for CRFs with Co-occurrence Statistics, International Journal of Computer Vision (IJCV) 103 (2) (2013) 213–225.

[124] J. D. Lafferty, A. McCallum, F. C. N. Pereira, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in: International Conference on Machine Learning (ICML), 2001.

[125] C. Sutton, A. McCallum, An Introduction to Conditional Random Fields, Foundations and Trends in Machine Learning 4 (4) (2012) 267–373.

[126] M. Pawan Kumar, Combinatorial and Convex Optimization for Probabilistic Models in Computer Vision, Ph.D. thesis, Oxford Brookes University (2008).

[127] Y. Boykov, M.-P. Jolly, Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images, in: IEEE International Conference on Computer Vision (ICCV), 2001.

[128] S. Kumar, M. Hebert, Discriminative Fields for Modeling Spatial Dependencies in Natural Images, in: Advances in Neural Information Processing Systems (NIPS), 2003.

[129] X. He, R. S. Zemel, M. A. Carreira-Perpinan, Multiscale Conditional Random Fields for Image Labeling, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004.

[130] L. Ladický, C. Russell, P. Kohli, P. H. S. Torr, Associative Hierarchical CRFs for Object Class Image Segmentation, in: IEEE International Conference on Computer Vision (ICCV), 2009.

[131] A. Quattoni, M. Collins, T. Darrell, Conditional Random Fields for Object Recognition, in: Advances in Neural Information Processing Systems (NIPS), 2004.

[132] L. Ladický, P. Sturgess, K. Alahari, C. Russell, P. H. S. Torr, What, Where & How Many? Combining Object Detectors and CRFs, in: European Conference on Computer Vision (ECCV), 2010.

[133] P. Krähenbühl, V. Koltun, Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, in: Advances in Neural Information Processing Systems (NIPS), 2011.

[134] J. Shotton, J. Winn, C. Rother, A. Criminisi, TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context, International Journal of Computer Vision (IJCV) 81 (1) (2009) 2–23.

[135] P. Krähenbühl, V. Koltun, Efficient Nonlocal Regularization for Optical Flow, in: European Conference on Computer Vision (ECCV), 2012.

[136] D. Sun, J. Wulff, E. B. Sudderth, H. Pfister, M. J. Black, A Fully-connected Layered Model of Foreground and Background Flow, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[137] N. D. Campbell, K. Subr, J. Kautz, Fully-Connected CRFs with Non-Parametric Pairwise Potential, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[138] A. P. Dawid, Applications of a General Propagation Algorithm for Probabilistic Expert Systems, Statistics and Computing 2 (1) (1992) 25–36.

[139] S. M. Aji, R. J. McEliece, The Generalized Distributive Law, IEEE Transactions on Information Theory 46 (2) (2000) 325–343.

[140] D. Schlesinger, B. Flach, Transforming an Arbitrary Minsum Problem into a Binary One, Tech. Rep. TUD-FI06-01, Dresden University of Technology (2006).

[141] J. Besag, On the Statistical Analysis of Dirty Pictures (with discussion), Journal of the Royal Statistical Society (Series B) 48 (3) (1986) 259–302.

[142] A. Blake, A. Zisserman, Visual Reconstruction, MIT Press, 1987.

[143] F. Tupin, H. Maitre, J.-F. Mangin, J.-M. Nicolas, E. Pechersky, Detection of Linear Features in SAR Images: Application to Road Network Extraction, IEEE Transactions on Geoscience and Remote Sensing 36 (2) (1998) 434–453.

[144] P. B. Chou, C. M. Brown, The Theory and Practice of Bayesian Image Labeling, International Journal of Computer Vision (IJCV) 4 (3) (1990) 185–210.

[145] P. B. Chou, P. R. Cooper, M. J. Swain, C. M. Brown, L. E. Wixson, Probabilistic Network Inference for Cooperative High and Low Level Vision, in: R. Chellappa, A. Jain (Eds.), Markov Random Fields: Theory and Applications, Academic Press, 1993, pp. 211–243.

[146] Y. Weiss, W. T. Freeman, On the Optimality of Solutions of the Max-product Belief-propagation Algorithm in Arbitrary Graphs, IEEE Transactions on Information Theory 47 (2) (2001) 736–744.

[147] P. F. Felzenszwalb, D. P. Huttenlocher, Efficient Belief Propagation for Early Vision, International Journal of Computer Vision (IJCV) 70 (1) (2006) 41–54.

[148] H. Ishikawa, D. Geiger, Segmentation by Grouping Junctions, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1998.

[149] E. Boros, P. L. Hammer, X. Sun, Network Flows and Minimization of Quadratic Pseudo-Boolean Functions, Tech. Rep. RRR 17-1991, RUTCOR Research Report (1991).

[150] V. Kolmogorov, C. Rother, Minimizing Nonsubmodular Functions with Graph Cuts - A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 29 (7) (2007) 1274–1279.

[151] E. Boros, P. L. Hammer, G. Tavares, Preprocessing of Unconstrained Quadratic Binary Optimization, Tech. Rep. RRR 10-2006, RUTCOR Research Report (2006).

[152] C. Rother, V. Kolmogorov, V. Lempitsky, M. Szummer, Optimizing Binary MRFs via Extended Roof Duality, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[153] N. Komodakis, G. Tziritas, Approximate Labeling via Graph Cuts Based on Linear Programming, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 29 (8) (2007) 1436–1453.

[154] N. Komodakis, G. Tziritas, N. Paragios, Fast, Approximately Optimal Solutions for Single and Dynamic MRFs, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[155] T. Werner, A Linear Programming Approach to Max-Sum Problem: A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 29 (7) (2007) 1165–1179.

[156] N. Komodakis, N. Paragios, G. Tziritas, MRF Energy Minimization and Beyond via Dual Decomposition, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 33 (3) (2011) 531–552.

[157] E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, M. Yannakakis, The complexity of Multiway Cuts (extended abstract), in: ACM Symposium on Theory of Computing (STOC), 1992.

[158] L. R. Ford, D. R. Fulkerson, Flows in Networks, Princeton University Press, 1962.

[159] A. V. Goldberg, R. E. Tarjan, A New Approach to the Maximum-flow Problem, Journal of the ACM (JACM) 35 (4) (1988) 921–940.

[160] V. V. Vazirani, Approximation Algorithms, Springer, 2001.

[161] E. Boros, P. L. Hammer, Pseudo-boolean Optimization, Discrete Applied Mathematics 123 (1-3) (2002) 155–225.

[162] S. Birchfield, C. Tomasi, Multiway Cut for Stereo and Motion with Slanted Surfaces, in: IEEE International Conference on Computer Vision (ICCV), 1999.

[163] Y. Boykov, M.-P. Jolly, Interactive Organ Segmentation Using Graph Cuts, in: International Conference, Medical Image Computing and Computer Assisted Intervention (MICCAI), 2000.

[164] D. Snow, P. Viola, R. Zabih, Exact Voxel Occupancy with Graph Cuts, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2000.

[165] Y. Boykov, O. Veksler, R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, in: International Conference on Computer Vision (ICCV), 1999.

[166] O. Veksler, Graph Cut Based Optimization for MRFs with Truncated Convex Priors, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[167] M. Pawan Kumar, O. Veksler, P. H. S. Torr, Improved Moves for Truncated Convex Models, Journal of Machine Learning Research 12 (2011) 31–67.

[168] O. Veksler, Multi-label Moves for MRFs with Truncated Convex Priors, International Journal of Computer Vision (IJCV) 98 (1) (2012) 1–14.

[169] O. Veksler, Dynamic Programming for Approximate Expansion Algorithm, in: European Conference on Computer Vision (ECCV), 2012.

[170] P. L. Hammer, P. Hansen, B. Simeone, Roof Duality, Complementation and Persistency in Quadratic 0-1 Optimization, Mathematical Programming 28 (2) (1984) 121–155.

[171] P. Kohli, A. Shekhovtsov, C. Rother, V. Kolmogorov, P. H. S. Torr, On Partial Optimality in Multi-label MRFs, in: International Conference on Machine Learning (ICML), 2008.

[172] V. Lempitsky, C. Rother, S. Roth, A. Blake, Fusion Moves for Markov Random Field Optimization, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 32 (8) (2010) 1392–1405.

[173] P. Kohli, P. H. S. Torr, Efficiently Solving Dynamic Markov Random Fields using Graph Cuts, in: IEEE International Conference on Computer Vision (ICCV), 2005.

[174] O. Juan, Y. Boykov, Active Graph Cuts, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006.

[175] K. Alahari, P. Kohli, P. H. S. Torr, Reduce, Reuse & Recycle: Efficiently Solving Multi-label MRFs, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[176] K. Alahari, P. Kohli, P. H. S. Torr, Dynamic Hybrid Algorithms for MAP Inference in Discrete MRFs, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 32 (10) (2010) 1846–1857.

[177] I. Kovtun, Partial Optimal Labeling Search for a NP-Hard Subclass of (max, +) Problems, in: DAGM Symposium, 2003.

[178] D. Batra, P. Kohli, Making the Right Moves: Guiding Alpha-Expansion using Local Primal-Dual Gaps, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[179] M. A. Paskin, Thin Junction Tree Filters for Simultaneous Localization and Mapping, in: International Joint Conference on Artificial Intelligence (IJCAI), 2003.

[180] P. F. Felzenszwalb, R. Zabih, Dynamic Programming and Graph Algorithms in Computer Vision, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 33 (4) (2011) 721–740.

[181] M. J. Wainwright, T. Jaakkola, A. Willsky, Tree Consistency and Bounds on the Performance of the Max-product Algorithm and Its Generalizations, Statistics and Computing 14 (2) (2004) 143–166.

[182] B. J. Frey, D. J. C. MacKay, A Revolution: Belief Propagation in Graphs with Cycles, in: Advances in Neural Information Processing Systems (NIPS), 1997.

[183] M. F. Tappen, W. T. Freeman, Comparison of Graph Cuts with Belief Propagation for Stereo, using Identical MRF Parameters, in: IEEE International Conference on Computer Vision (ICCV), 2003.

[184] M. Pawan Kumar, P. H. S. Torr, Fast Memory-efficient Generalized Belief Propagation, in: European Conference on Computer Vision (ECCV), 2006.

[185] K. Petersen, J. Fehr, H. Burkhardt, Fast Generalized Belief Propagation for MAP Estimation on 2D and 3D Grid-like Markov Random Fields, in: DAGM Symposium, 2008.

[186] G. Borgefors, Distance Transformations in Digital Images, Computer Vision, Graphics, and Image Processing 34 (3) (1986) 344–371.

[187] S. Alchatzidis, A. Sotiras, N. Paragios, Efficient Parallel Message Computation for MAP Inference, in: IEEE International Conference on Computer Vision (ICCV), 2011.

[188] U. Kjæ rulff, Inference in Bayesian Networks using Nested Junction Trees, in: M. I. Jordan (Ed.), Learning in Graphical Models, MIT Press, 1999, pp. 51–74.

[189] M. J. Wainwright, M. I. Jordan, Graphical Models, Exponential Families, and Variational Inference, Foundations and Trends in Machine Learning 1 (1-2) (2008) 1–305.

[190] J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnorr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, J. Lellmann, N. Komodakis, C. Rother, A Comparative Study of Modern Inference Techniques for Discrete Energy Minimization Problems, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[191] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[192] C. Yanover, T. Meltzer, Y. Weiss, Linear Programming Relaxations and Belief Propagation - An Empirical Study, The Journal of Machine Learning Research 7 (2006) 1887–1907.

[193] V. Kolmogorov, M. J. Wainwright, On the Optimality of Tree-reweighted Max-product Message-passing, in: Conference on Uncertainty in Artificial Intelligence (UAI), 2005.

[194] A. Globerson, T. Jaakkola, Fixing Max-Product: Convergent Message Passing Algorithms for MAP LP-Relaxations, in: Advances in Neural Information Processing Systems (NIPS), 2007.

[195] V. A. Kovalevsky, V. K. Koval, A Diffusion Algorithm for Decreasing Energy of Max-sum Labeling Problem, Tech. rep., Glushkov Institute Of Cybernetics, Kiev, USSR (1975).

[196] V. K. Koval, M. I. Schlesinger, Dvumernoe programmirovanie v zadachakh analiza izobrazheniy (Two-dimensional programming in image analysis problems), USSR Academy of Science, Automatics and Telemechanics 8 (1976) 149–168.

[197] D. Sontag, T. Jaakkola, New Outer Bounds on the Marginal Polytope, in: Advances in Neural Information Processing Systems (NIPS), 2007.

[198] D. Sontag, T. Meltzer, A. Globerson, T. Jaakkola, Y. Weiss, Tightening LP Relaxations for MAP using Message Passing, in: Conference on Uncertainty in Artificial Intelligence (UAI), 2008.

[199] N. Komodakis, N. Paragios, Beyond Loose LP-relaxations: Optimizing MRFs by Repairing Cycles, in: European Conference on Computer Vision (ECCV), 2008.

[200] T. Werner, Revisiting the Linear Programming Relaxation Approach to Gibbs Energy Minimization and Weighted Constraint Satisfaction, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 32 (8) (2010) 1474–1488.

[201] D. Batra, S. Nowozin, P. Kohli, Tighter Relaxations for MAP-MRF Inference: A Local Primal-Dual Gap based Separation Algorithm, Journal of Machine Learning Research - Proceedings Track 15 (2011) 146–154.

[202] D. P. Bertsekas, Nonlinear Programming (Second Edition), Athena Scientific, 1999.

[203] L. Torresani, V. Kolmogorov, C. Rother, Feature Correspondence via Graph Matching: Models and Global Optimization, in: European Conference on Computer Vision (ECCV), 2008.

[204] S. Vicente, V. Kolmogorov, C. Rother, Joint Optimization of Segmentation and Appearance Models, in: IEEE International Conference on Computer Vision (ICCV), 2009.

[205] V. Jojic, S. Gould, D. Koller, Accelerated Dual Decomposition for MAP inference, in: International Conference on Machine Learning (ICML), 2010.

[206] P. Strandmark, F. Kahl, Parallel and Distributed Graph Cuts by Dual Decomposition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[207] S. Ramalingam, P. Kohli, K. Alahari, P. H. S. Torr, Exact Inference in Multi-label CRFs with Higher Order Cliques, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[208] X. Lan, S. Roth, D. P. Huttenlocher, M. J. Black, Efficient Belief Propagation with Learned Higher-Order Markov Random Fields, in: European Conference on Computer Vision (ECCV), 2006.

[209] B. Potetz, T. S. Lee, Efficient Belief Propagation for Higher-order Cliques using Linear Constraint Nodes, Computer Vision and Image Understanding (CVIU) 112 (1) (2008) 39–54.

[210] H. Ishikawa, Higher-order Clique Reduction in Binary Graph Cut, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[211] A. Fix, A. Gruber, E. Boros, R. Zabih, A Graph Cut Algorithm for Higher-order Markov Random Fields, in: IEEE International Conference on Computer Vision (ICCV), 2011.

[212] I. G. Rosenberg, Reduction of Bivalent Maximization to the Quadratic Case, Cahiers du Centre d'etudes de Recherche Operationnelle 17 (1975) 71–74.

[213] A. M. Ali, A. A. Farag, G. L. Gimel'farb, Optimizing Binary MRFs with Higher Order Cliques, in: European Conference on Computer Vision (ECCV), 2008.

[214] D. Freedman, P. Drineas, Energy Minimization via Graph Cuts: Settling What is Possible, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[215] H. Ishikawa, Transformation of General Binary MRF Minimization to the First Order Case, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 33 (6) (2011) 1234–1249.

[216] A. C. Gallagher, D. Batra, D. Parikh, Inference for Order Reduction in Markov Random Fields, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[217] A. Delong, L. Gorelick, O. Veksler, Y. Boykov, Minimizing Energies with Hierarchical Costs, International Journal of Computer Vision (IJCV) 100 (1) (2012) 38–58.

[218] B. Potetz, Efficient Belief Propagation for Vision Using Linear Constraint Nodes, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[219] D. Tarlow, I. E. Givoni, R. S. Zemel, HOP-MAP: Efficient Message Passing with High Order Potentials, in: International Conference on Artificial Intelligence and Statistics (AISTATS), 2010.

[220] J. J. Mcauley, T. S. Caetano, Faster Algorithms for Max-Product Message-Passing, Journal of Machine Learning Research 12 (2011) 1349–1388.

[221] P. F. Felzenszwalb, J. J. Mcauley, Fast Inference with Min-Sum Matrix Product, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 33 (12) (2011) 2549–2554.

[222] T. Werner, High-arity Interactions, Polyhedral Relaxations, and Cutting Plane Algorithm for Soft Constraint Optimisation (MAP-MRF), in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[223] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, N. Paragios, A Generic Local Deformation Model for Shape Registration, Tech. Rep. RR-7676, INRIA (July 2011).

[224] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, N. Paragios, Dense Non-rigid Surface Registration using High-order Graph Matching, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[225] C. Rother, P. Kohli, W. Feng, J. Jia, Minimizing Sparse Higher Order Energy Functions of Discrete Variables, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[226] P. Kohli, M. Pawan Kumar, Energy Minimization for Linear Envelope MRFs, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[227] A. Delong, O. Veksler, A. Osokin, Y. Boykov, Minimizing Sparse High-Order Energies by Submodular Vertex-Cover, in: Advances in Neural Information Processing Systems (NIPS), 2012.

[228] Y. Zeng, C. Wang, S. Soatto, S.-T. Yau, Nonlinearly Constrained MRFs: Exploring the Intrinsic Dimensions of Higher-Order Cliques, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[229] B. Taskar, C. Guestrin, D. Koller, Max-margin Markov Networks, in: Advances in Neural Information Processing Systems (NIPS), 2003.

[230] D. Munoz, J. A. D. Bagnell, N. Vandapel, M. Hebert, Contextual Classification with Functional Max-Margin Markov Networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[231] M. Szummer, P. Kohli, D. Hoiem, Learning CRFs using Graph Cuts, in: European Conference on Computer Vision (ECCV), 2008.

[232] S. Gould, Max-margin Learning for Lower Linear Envelope Potentials in Binary Markov Random Fields, in: International Conference on Machine Learning (ICML), 2011.

[233] D. Tarlow, R. S. Zemel, Structured Output Learning with High Order Loss Functions, in: International Conference on Artificial Intelligence and Statistics (AISTATS), 2012.

[234] P. Pletscher, P. Kohli, Learning Low-order Models for Enforcing High-order Statistics, in: International Conference on Artificial Intelligence and Statistics (AISTATS), 2012.

[235] T. Finley, T. Joachims, Training Structural SVMs when Exact Inference is Intractable, in: International Conference on Machine Learning (ICML), 2008.

[236] Y. Li, D. P. Huttenlocher, Learning for Stereo Vision using the Structured Support Vector Machine, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[237] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, A. Ng, Discriminative Learning of Markov Random Fields for Segmentation of 3D Scan Data, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[238] N. Komodakis, Efficient Training for Pairwise or Higher Order CRFs via Dual Decomposition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[239] C.-N. J. Yu, T. Joachims, Learning Structural SVMs with Latent Variables, in: International Conference on Machine Learning (ICML), 2009.

[240] N. Komodakis, Learning to Cluster Using High Order Graphical Models with Latent Variables, in: IEEE International Conference on Computer Vision (ICCV), 2011.

[241] M. Pawan Kumar, B. Packer, D. Koller, Modeling Latent Variable Uncertainty for Loss-based Learning, in: International Conference on Machine Learning (ICML), 2012.

[242] K. G. G. Samuel, M. F. Tappen, Learning Optimized MAP Estimates in Continuously-valued MRF Models, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.