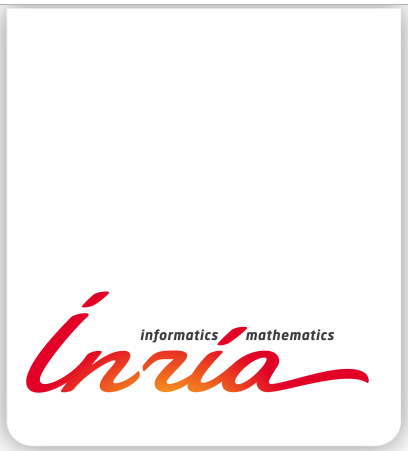


Markov Random Fields in Vision Perception: A Survey

Chaohui Wang, Nikos Paragios

RESEARCH REPORT
N° 7945
September 2012
Project-Teams GALEN

ISSN 0249-6399 ISRN INRIA/RR--7945--FR+ENG



Markov Random Fields in Vision Perception: A Survey

Chaohui Wang*†, Nikos Paragios*†

Project-Teams GALEN

Research Report n° 7945 — September 2012 — 42 pages

Abstract: In this paper, we present a comprehensive survey of Markov Random Fields (MRFs) in computer vision, with respect to both the modeling and the inference. MRFs were introduced into the computer vision field about two decades ago, while they started to become a ubiquitous tool for solving visual perception problems at the turn of the millennium following the emergence of efficient inference methods. During the past decade, different MRF models as well as inference methods - in particular those based on discrete optimization - have been developed towards addressing numerous vision problems of low, mid and high level. While most of the literature concerns pairwise MRFs, during recent years, we have also witnessed significant progress on higher-order MRFs, which substantially enhances the expressiveness of graph-based models and enlarges the extent of solvable problems. We hope that this survey will provide a compact and informative summary of the main literature in this research topic.

Key-words: Markov Random Fields, Graphical Models, MRFs, Graph-based Methods, MAP Inference, Discrete Optimization

* Center for Visual Computing, École Centrale Paris, Châtenay-Malabry, France
† Equipe GALEN, INRIA Saclay - Île de France, Orsay, France

RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

Parc Orsay Université
4 rue Jacques Monod
91893 Orsay Cedex

Résumé : Dans cet article, nous présentons un panorama approfondi des champs de Markov aléatoires (MRFs) dans le cadre de la vision par ordinateur, et ce autant du point de vue de la modélisation que de l'inférence. Les MRFs ont été introduits dans le domaine de la vision par ordinateur il y a environ deux décennies, alors qu'ils commençaient à devenir un outil omniprésent pour résoudre les problèmes de perception visuelle à la suite de l'apparition de méthodes efficaces d'inférence. Au cours de la dernière décennie, les différents modèles de MRFs ainsi que les méthodes d'inférence - en particulier celles basées sur l'optimisation discrète, ont été mis en oeuvre pour résoudre de nombreux problèmes de vision de bas, milieu et haut niveaux. Alors que la plupart de la littérature concerne les MRFs d'ordre deux, nous avons également assisté au cours des dernières années à des progrès significatifs sur les MRFs d'ordre supérieur, ce qui améliore sensiblement l'expressivité des modèles à base de graphes et élargit le champs d'application de ces méthodes. Nous espérons que cette étude bibliographique fournira un résumé compact et informatif sur la littérature principale concernant ce sujet de recherche.

Mots-clés : Champs de Markov Aléatoires, Modèles graphiques, MRFs, Inférence MAP, Optimisation discrète

1 Introduction

The goal of computer vision is to enable the machine to understand the world - often called *visual perception* - through processing of digital signals. Such an understanding for the machine is done by extracting useful information from the signals and performing complex reasoning. Mathematically, let $\mathbf{I}$ denote the observed data and $\mathbf{x}$ a latent parameter vector of interest that corresponds to a mathematical answer to the visual perception problem. Visual perception can be formulated mathematically as finding a mapping from $\mathbf{I}$ to $\mathbf{x}$, which is essentially an *inverse problem* [177]. Mathematical methods usually model such a mapping through an optimization problem as follows:

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

where the energy (or cost, objective) function $E(\mathbf{x};\mathbf{I})$ can be regarded as a quality measure of a parameter configuration $\mathbf{x}$ in the solution space, given the observed images $\mathbf{I}$. Hence, visual perception involves two main tasks: *modeling* and *inference/optimization*. The former 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 energy function $E(\mathbf{x};\mathbf{I})$ which can correctly measure the adequacy between $\mathbf{x}$ and $\mathbf{I}$. The latter has to search for the configuration of $\mathbf{x}$ producing the optimum of the energy function where the solution of the original problem lies. The main difficulties in the modeling are due to the fact that most of the vision problems are inverse and 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 well 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 modeling is the tractability of the inference task. Such a viewpoint impacts the quality of the obtained optima and introduce 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, due to the ability to model soft contextual constraints between variables and enormous development of inference methods, 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 algorithms. Generally speaking, they have the following major useful properties that one can benefit during the algorithm design. First, graphical models refer to 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 graphical models provides a simple way to visualize the structure of a model and facilitates the choice and design of the model. Third, the factorization of the joint probability over a graph could produce inference problems that can be solved in a computational efficient manner. In particular, development of inference methods based on discrete optimization enhance the potential of discrete MRFs and enlarge significantly the set of visual perception problems on which they can be applied. Last but not least, the probabilistic side of graphical models leads to potential advantages in terms of parameter learning (*e.g.*, [152, 160]) and uncertainty analysis (*e.g.*, [96, 62]) over classic variational methods

[183, 40], due to the introduction of probability explanation to the solution [177]. 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 methods - in particular those based on discrete optimization - have been developed towards addressing numerous vision problems of low, mid and high-level. 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 enlarges the extent of solvable problems. We believe that a compact and informative summary of the main literature in this research topic will be very helpful for related researchers to get a global view and hence better understanding of such an important tool. To this end, we present in this paper a compressive survey of MRFs in computer vision, with respect to both the modeling and the inference.

The survey is structured from basic notions to most recent important works on them, aiming to make it as self-complete as possible and beneficial to readers of different backgrounds. The remainder of this paper is organized as follows. section 2) introduces basic knowledge on graphical models. In what follows, the paper focuses on MRFs: in section 3, different important subclasses of MRFs are discussed as well as their important applications in visual perception, and then representative techniques for the MAP inference in discrete MRFs are presents in section 4. Finally, we conclude the survey in section 5.

2 Preliminaries: Graphical Models

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 of which involves 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 that is called *Factor Graph*, in particular for better visualizing potential functions and performing inference in higher-order models. We will proceed with a formal brief presentation of each model in the reminder of this section. We suggest the reader being interested for a larger and more in depth overview the following publications [123, 16, 79, 97].

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 ($i \in \mathcal{V}$) contained in the model, 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 for all individual variables, *i.e.*, $\mathcal{X} = \prod_{i \in \mathcal{V}} \mathcal{X}_i$.

For the purposes of simplification and concreteness, we use "probability distribution" 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, for the purpose of convenience, we often use "node" to refer to the corresponding random variable in cases where there is no ambiguity.

2.2 Bayesian Networks (Directed Graphical Models)

A *Bayesian Network (BN)* has the structure of a directed acyclic graph (DAG) $\mathcal{G}$ where the edges in $\mathcal{E}$ are directed and no directed cycle exists (*e.g.*, Fig. 1(a)), and holds the following local independence assumptions (referred to as *local Markov property*) which impose that every node is independent of its non-descendant nodes[1] given all its parents:

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

where $\mathcal{A}_i$ and π_i denotes the set of non-descendant nodes and the set of parents for a node i in the graph $\mathcal{G}$, respectively, and $X_i \perp X_j | X_k$ denotes the statement that X_i and X_j are independent given X_k. The associated family of joint probability distributions are those satisfying the local independences in Eq. 2, and can be factorized into the following form according to $\mathcal{G}$:

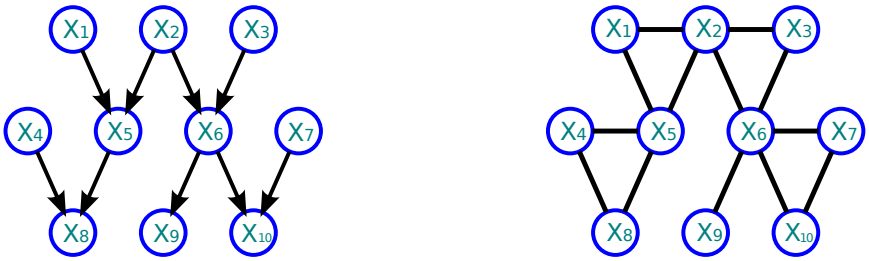
$$p(\mathbf{x}) = \prod_{i \in \mathcal{V}} p(x_i | x_{\pi_i}) \tag{3}$$

where $p(x_i|x_{\pi_i})$ denotes local conditional probability distribution (CPD) of x_i given the states x_{π_i} of the parents. It should be noted that any distribution with the factorized form in Eq. 3 satisfies the local independences in Eq. 2.

All conditional independences (referred to as *global Markov property*) implied within the structure of BNs, including the local independences of Eq. 2, can be identified by checking *d-separation* properties of the corresponding graph $\mathcal{G}$ [139]. This can be performed using an intuitive and handy method: *Bayes ball algorithm* [56, 163]. Let $\mathcal{I}(\mathcal{G})$ denote the set of such conditional independences. Note that the global Markov property and the local Markov property are equivalent in BNs. Hence, if a distribution can be factorized over $\mathcal{G}$, it must satisfy all the conditional independences in $\mathcal{I}(\mathcal{G})$. On the other hand, we should also note that an instance of distribution that can be factorized over $\mathcal{G}$ may satisfy more independences than those in $\mathcal{I}(\mathcal{G})$. Nevertheless, such instances are very "few" in the sense that they have measure zero in the space of CPD parameterizations, *e.g.*, a slight perturbation of the local CPDs will almost certainly eliminate these "extra" independences [97].

BNs are usually used to model causal relationships [140] between random variables and have been applied in many fields such as artificial intelligence, computer vision, automatic control, information engineering, *etc.* In computer vision, Hidden Markov Models (HMM) [146] and Kalman Filters [81, 57], which are well-known subsets of BNs, provide a common way to model temporal relations and has been employed to deal with object tracking [182, 203], denoising [85, 149], motion analysis [71, 65], sign language recognition [170, 130], *etc.* Besides, neural networks [15], another special type of BNs, provide an important machine learning method to deal with vision problems [38]. Other vision applications include for example [134] and [210],

[1] For a node $i \in \mathcal{V}$, its non-descendant nodes consist of the nodes $j \in \mathcal{V} - \{i\}$ such that there is no directed path from i to j.



(a) Bayesian Network (b) Markov Random Filed

Figure 1: Examples of Bayesian Network and Markov Random Filed. Note that the directed graph in (a) can be transformed into the undirected graph in (b) by *moralization* process [79].

where dynamic BNs have been used to perform gesture/speech recognition and facial expression understanding, respectively.

2.3 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(b)). Furthermore, such a paradigm inherits the following local independence assumptions (also referred to as *local Markov property*):

$$\forall\, i \in \mathcal{V}, X_i \perp X_{\mathcal{V}-\{i\}} | X_{\mathcal{N}_i} \tag{4}$$

which impose that a node is independent of any other node given all its neighbors. In such a context, $\mathcal{N}_i = \{j | \{i,j\} \in \mathcal{E}\}$ denotes the set of neighbors of node i in the graph $\mathcal{G}$. An important notion in MRFs is *clique*, which is defined as a full-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. 4). According to Hammersley-Clifford theorem [67, 12], they are *Gibbs distributions* which can be factorized into the following form according to $\mathcal{G}$:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(x_c) \tag{5}$$

where Z is the normalizing factor (also known as the *partition function*), $\psi_c(x_c)$ denotes the *potential function* of a clique c which is a positive real-valued function on the possible configuration x_c of the clique c, and $\mathcal{C}$ 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. 5 satisfies the local Markov property in Eq. 4.

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 $X_{\mathcal{V}_1} \perp X_{\mathcal{V}_2} | X_{\mathcal{V}_3}$. Let $\mathcal{I}(\mathcal{G})$ denote the set

[2]Note that any quantities defined on a non-maximal clique can always be redefined on the corresponding maximal clique, and thus $\mathcal{C}$ 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 usually leads to computational efficiency in practice. Hence, without loss of generality, we do not assume that $\mathcal{C}$ consist of only maximal cliques in this survey.

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 $\mathcal{V}_1 - \mathcal{V}_3$ and any node in $\mathcal{V}_2 - \mathcal{V}_3$. This problem can be solved using standard search algorithms such as breadth-first search (BFS) [34]. 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. 5 according to $\mathcal{G}$, then it satisfies all the conditional independences in $\mathcal{I}(\mathcal{G})$. Similar to Bayesian Network, an instance of distribution that can be factorized over $\mathcal{G}$, may satisfies more independences than those in $\mathcal{I}(\mathcal{G})$.

MRFs provide a principled probabilistic framework to model vision problems, thanks to their ability to model soft contextual constraints between random variables [127]. The adoption of such constraints is important in vision problems, since the image and/or scene modeling 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:

$$\mathbf{x}^{\text{opt}} = \arg\max_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x}) \tag{6}$$

Since the potential functions are restricted to positive here, let us define clique energy θ_c as a real function on a clique c ($c \in \mathcal{C}$):

$$\theta_c(x_c) = -\log \psi_c(x_c) \tag{7}$$

Due to the one-to-one mapping between θ_c and ψ_c, we also call θ_c *potential function* (or *clique potential*) on clique c in the remaining of this survey, which leads to a more convenient representation of the joint distribution $p(\mathbf{x})$:

$$p(\mathbf{x}) = \frac{1}{Z} \exp\{-E(\mathbf{x})\} \tag{8}$$

where $E(\mathbf{x})$ denotes the *energy* of the MRF and is defined as a sum of potential functions on the cliques:

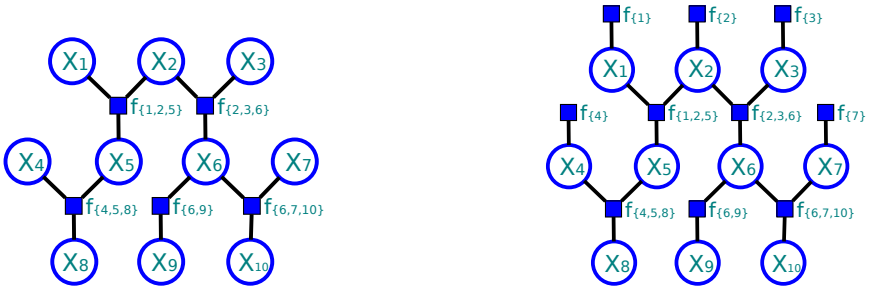
$$E(\mathbf{x}) = \sum_{c\in\mathcal{C}} \theta_c(x_c) \tag{9}$$

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 (*i.e.*, the maximization of $p(\mathbf{x})$ in Eq. 6) is equivalent to the minimization of $E(\mathbf{x})$ as follows:

$$\mathbf{x}^{\text{opt}} = \arg\min_{\mathbf{x}\in\mathcal{X}} E(\mathbf{x}) \tag{10}$$

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 optimization/inference algorithms using discrete optimization theories and techniques (*e.g.*, [28, 73, 102, 193, 95, 98, 110, 136, 103]), which have been successfully employed to efficiently solve many vision problems using MRF-based methods (*e.g.*, [101, 61, 92, 178, 23]). Due to the advantages regarding both the modeling and the inference

[3]We should note that *continuous MRFs* have also been used in the literature (*e.g.*, [72, 167, 172]). An important subset of continuous MRFs that has been well studied is *Gaussian MRF* [159].



(a) (b)

Figure 2: Examples of Factor Graphs. Note that both of the Bayesian Network in Fig. 1(a) and the Markov Random Filed in Fig. 1(b) can be represented by the two factor graphs above. Nevertheless, the factor graph in (b) contains factors corresponding to non-maximal cliques.

as discussed above, 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 their MAP inference methods in section 4.

2.4 Factor Graphs

Factor graph [52, 114] is a unified representation for both BNs and MRFs, which uses additional nodes, named *factor nodes*[4], to explicitly describe the factorization of the joint distribution in the graph. More specifically, a set $\mathcal{F}$ of factor nodes are introduced into the graph, corresponding each to an objective function term defined on a subset of usual nodes. Each factor encodes a local conditional probability distribution defined on a usual node and its parents in cases of BNs (see Eq. 3), while it encodes a potential function defined on a clique in cases of MRFs (see Eq. 5 or Eq. 9). The associated joint probability is a product of factors:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{f \in \mathcal{F}} \phi_f(x_f) \tag{11}$$

where the normalizing factor Z is equal to 1 for BNs. Similar to MRFs, we can define the energy of the factor graph as:

$$E(\mathbf{x}) = \sum_{f \in \mathcal{F}} \theta_f(x_f) \tag{12}$$

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. 2 shows two examples of factor graphs which provide two different possible representations for both the BN in Fig. 1(a) and the MRF in Fig. 1(b).

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. 2(b)). The same objective can be hardly met using the usual graphical representation of MRFs. Computational inference

[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.

is another strength of factor graphs representations. The *sum-product* and *min-sum* (or: *max-product*[5]) algorithms in the factor graph [114, 16] generalize the classic counterparts [139, 206] in the sense that the order of factors can be greater than two, which will be presented in section 4.2. Furthermore, since an MRF with loops may has no loop in its corresponding factor graph (*e.g.*, see the MRF in Fig. 1(b) and the factor graphs in Fig. 2 (a-b)), in such cases the *min-sum* algorithm in the factor graph can perform the MAP inference exactly with polynomial complexity. Such factor graphs without loop (*e.g.*, Fig. 2 (a-b)) 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 subset is *Conditional Random Fields* (CRFs). Below, we present these three typical subsets of MRFs 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 *singleton potentials* (also known as *unary 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:

$$E(\mathbf{x}) = \sum_{i\in\mathcal{V}} \theta_i(x_i) + \sum_{\{i,j\}\in\mathcal{E}} \theta_{ij}(x_{ij}) \tag{13}$$

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. On top of that, their use was spread also due to the fact 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.*, [58, 155, 42, 23]) and the efficient inference in pairwise MRFs (*e.g.*, [28, 193, 98, 95, 106]). Two most typical graph structures used in computer vision are *grid-like structures* (*e.g.*, Fig. 3) and *part-based structures* (*e.g.*, Fig. 4). Grid-like structures provide a natural and reasonable representation for images, while part-based structures are often associated with deformable and/or articulated objects.

3.1.1 Grid-like Models

Pairwise MRFs of *grid-like structures* (Fig. 3) have been widely used in computer vision to deal with a large number of important problems, such as image denoising/restoration (*e.g.*, [58, 64, 30]), super-resolution (*e.g.*, [51, 50, 147]), stereo vision/multi-view reconstruction (*e.g.*, [157, 101, 191]), optical flow and motion analysis (*e.g.*, [70, 158, 62, 128]), image registration and matching (*e.g.*, [175, 61, 164]), segmentation (*e.g.*, [25, 155, 23, 168]) and over-segmentation (*e.g.*, [131, 188, 211]).

[5]The *max-product* algorithm is to maximize the probability $p(\mathbf{x})$ which is a product of local functions (Eq. 11), while the *min-sum* algorithm is to minimize the corresponding energy which is a sum of local energy functions (Eq. 12). They are essentially the same algorithm.



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

Figure 3: Examples of MRFs with Grid-like Structures

In this context, the nodes of an MRF correspond to the lattice of pixels[6] and 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[7] (*e.g.*, an index denoting the segment that the corresponding pixel belongs to 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 singleton potentials $\theta_i(\cdot)$, whose definition is specific to the considered applications (*e.g.*, for image denoising, such singleton 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 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:

$$\theta_{ij}(x_{ij}) = \rho(x_i - x_j) \tag{14}$$

where $\rho(\cdot)$ is usually an even and non-decreasing function. In computer vision, common choices (Eq. 15) for $\rho(\cdot)$ are *(generalized) Potts model*[8] [144, 26], *truncated absolute distance* and *truncated quadratic*, which are typical *discontinuity preserving* penalties:

$$\rho(x_i - x_j) = \begin{cases} w_{ij} \cdot (1 - \delta(x_i - x_j)) & \text{(Potts models)} \\ \min(K_{ij}, |x_i - x_j|) & \text{(truncated absolute distance)} \\ \min(K_{ij}, (x_i - x_j)^2) & \text{(truncated quadratic)} \end{cases} \tag{15}$$

where $w_{ij} \geq 0$ is a weight coefficient[9] for the penalities, 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 [181, 124]. Such discontinuity preserving terms reduce the risk of over-smoothing, which is an advantage compared with Total Variation (TV) regularizations [31] that are often used in variational methods [183, 40]. Last, it should also be mentioned that pairwise potentials in such grid-like MRFs can also be used to encode other contextual constraints, such as the star shape prior in [187] and the layer constraint in [131].

[6] Other homogeneously distributed units such as 3D voxels and control points [61] can also be considered in such MRFs.

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

[8] Note that *Ising model* [77, 58] is a particular case of *Potts model* where each node has two possible states.

[9] w_{ij} is a constant for all pairs $\{i,j\}$ of nodes in the original Potts model in [144].



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

Figure 4: Example of MRFs with Pictorial Structures (The original image used in (a) is from *HumanEva-I* database: `http://vision.cs.brown.edu/humaneva/`.)

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.*, [54, 7, 126]), 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.*, [82, 6, 209]), where we aim to infer the configuration of each vertex or facet on the surface. In these cases, the node of MRF models can be used to model the superpixel, vertex or face, nevertheless, the topology could be a less regular grid.

3.1.2 Part-based Models

MRFs of *pictorial structures* (Fig. 4) 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 the interactions such as kinematic constraints between the corresponding pair of components. In [42], *Pictorial model* [46] was introduced into computer vision to deal with pose recognition of human body and face. In this work, a tree-like MRF (see Fig. 4) was employed to model the spring-like prior between pairs of components through pairwise potentials, while the data likelihood is encoded in the singleton 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 [10, 34] (*i.e.*, min-sum belief propagation [139, 206, 16]). This work has gained a lot of attention in computer vision and the proposed 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 [167, 166, 39, 5], hand [174, 173] and other objects [138, 41]. In [138], part-based model of [42] was extended 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. Continuous MRFs of pictorial structures were proposed in [167] and [174] to deal with body and/or hand tracking, where nonparametric belief propagation algorithms [72, 172] were employed to perform inference. In the subsequent papers [166, 173], 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. More recently, a rigorous visibility modeling in graphical models was proposed in [195], where image segmentation, multi-object tracking and depth ordering 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 boundary of the shape. 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* [69]. The characterization of shape prior 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 has been employed to efficiently solving problems related to the inference of the shape model such as knowledge-based object segmentation (*e.g.*, [162, 14]). Follow this line of research, recently [204] 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 pair 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 solving them. 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 diverse vision models and inference methods have been proposed.

3.2 Higher-order MRF Models

Higher-order MRFs (also referred to as *high-order MRFs*) 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 advatages 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 between random variables and richer interactions between them, in order to improve the performance of the method. One can cite for example the higher-order MRF model proposed in [151, 153] 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 [151, 153]. $\mathcal{P}^n$ *Potts model* was proposed in [90, 91], which consists of a strict generalization of the generalized Potts model [26] (see Eq. 15). It considers a similar interaction between n nodes (instead of between two nodes) and its performance was demonstrated in image segmentation being a natural application domain of such a model. In [87, 88], $\mathcal{P}^n$ Potts model was further enriched towards a *robust* $\mathcal{P}^n$ *model*, which produced better segmentation performance. Higher-order smoothness priors were used in [202] to solve stereo reconstruction problems. Other types of higher-order pattern potentials were also considered in [105] 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 the cases where measures that intrinsically involve more than two variables are present. 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") [84]. In [4], 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 [116]. In the optical flow formulation proposed in [60], higher-order potentials were used to encode angle deviation prior, non-affine motion prior as well as the data likelihood. 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 [196, 197]. Such approaches avoid explicit estimation of the global transformation such as 3D pose (translation, rotation and scaling) and 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. One can cite for example [189] and [132] where global connectivity priors (*e.g.*, foreground segment must be connected) were used to enforce the connectedness of the resulting labels for binary image segmentation, [37] where 'label costs" [212] was introduced into graph-based segmentation formulation to deal with unsupervised image segmentation, and [118, 119] which proposed to incorporate "object co-occurrence statistics" in Conditional Random Field (CRF) models to object class image segmentation.

3.3 Conditional Random Fields

A Conditional Random Field (CRF) [121, 176] 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). It can be viewed as an MRF which is globally conditioned on the observed data $\mathbf{D}$. 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:

$$\forall\, i \in \mathcal{V}, X_i \perp X_{\mathcal{V}-\{i\}} | \{X_{\mathcal{N}_i}, \mathbf{D}\} \tag{16}$$

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:

$$p(\mathbf{x}|\mathbf{D}) = \frac{1}{Z(\mathbf{D})} \exp\{-E(\mathbf{x};\mathbf{D})\} \tag{17}$$

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

$$E(\mathbf{x};\mathbf{D}) = \sum_{c\in\mathcal{C}} \theta_c(x_c;\mathbf{D}) \tag{18}$$

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 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[10] 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

[10]Like [135], we use the term *generative MRFs* to distinguish the usual MRFs from CRFs.

adopted for example in binary image segmentation [24]. Nevertheless, 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, for example, grid-like CRFs was introduced in [115] to model spatial dependencies in the image, an approach that outperformed the classic MRF model [58] in the image restoration experiments. A multi-scale CRF model was proposed in [68] for object class image segmentation, and a more sophisticated model named "associative hierarchical CRFs" were proposed in [117] to solve the same problem. Following that, in [120], object detectors and CRFs were combined within a CRF model which can be solved efficiently, so as to jointly estimate the class category, location, and segmentation of objects/regions from 2D images. CRFs has been also applied for object recognition. For example, a discriminative part-based approach was proposed in [145] to recognize objects based on a tree-structured CRF. Very recently, [113] 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. 3) [165] and *robust* $\mathcal{P}^n$ *model* [88].

4 MAP Inference Methods for Discrete MRFs

An essential problem regarding the application of MRF models is how to infer the configuration for each of the nodes contained in an MRF. The MAP inference (*i.e.*, Eq. 6) in discrete MRFs, which boils down to an energy minimization problem as shown in Eq. 10. Such a combinatorial problem is known to be NP-hard in general [28, 102], except for some particular cases such as MRFs of bounded tree-width [36, 1, 79] (*e.g.*, tree-structured MRFs [139]) and pairwise MRFs with submodular energy [102, 161].

The most well-known early (before the 1990s) algorithms for optimizing the MRF energy were *iterated conditional modes* (ICM) [13], *simulated annealing* methods (*e.g.*, [58, 17, 185]) and *highest confidence first (HCF)* [33, 32]. While being computational efficient methods, 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.*, [51, 198, 43]) and *graph cuts* techniques (*e.g.*, [64, 157, 26, 76, 28]), provided powerful alternatives to the aforementioned methods from both computational and theoretical viewpoint and have been used to solve numerous visual perception problems (*e.g.*, [51, 175, 64, 76, 101, 25, 155]). Since then, the MRF optimization is experiencing a renaissance, and more and more researchers have been working on it. For the most recent MRF optimization techniques, one can cite for example *QPBO* techniques (*e.g.*, [19, 99, 20, 156]), LP primal-dual algorithms (*e.g.*, [108, 109, 110]) as well as dual methods (*e.g.*, [193, 98, 109, 199]).

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, we will describe the *min-sum belief propagation* algorithm in factor tree and also show its extensions towards dealing with an arbitrary pairwise MRF. Following that, we review in section 4.3 recent developed dual methods for pairwise MRFs, in particular the *tree-reweighted message passing* methods (*e.g.*, [193, 98]) and the

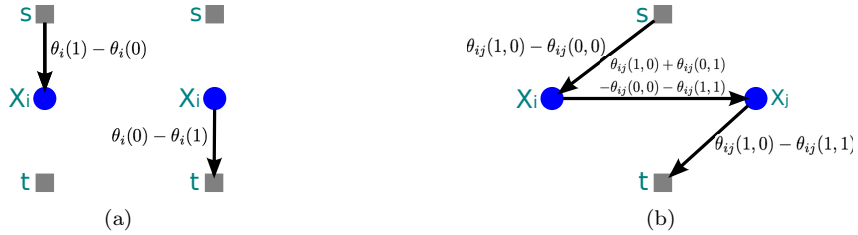


Figure 5: Examples of *s-t Graph* Construction for Binary Graph Cuts [102]. (a) Graphs for the singleton potential defined on a node i. The left one is for the cases where $\theta_i(0) < \theta_i(1)$ and the right one is for the cases where $\theta_i(0) \geq \theta_i(1)$; (b) Graph for the pairwise potential defined on an edge $\{i, j\}$ where $\theta_{ij}(1,0) > \theta_{ij}(0,0)$ and $\theta_{ij}(1,0) > \theta_{ij}(1,1)$. Note that $\theta_{ij}(1,0) + \theta_{ij}(0,1) - \theta_{ij}(0,0) - \theta_{ij}(1,1) > 0$ holds when the energy is submodular.

dual-decomposition approaches (*e.g.*, [109, 107]). 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 various vision problems (*e.g.*, [64, 76, 155, 101, 23, 92]).

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*[11], see examples in Fig. 5) 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 $C(S,T)$ (Eq. 19)) 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 or differ by a constant difference from the energy of the MRF with the corresponding configuration[12] $\mathbf{x}$.

$$C(S,T) = \sum_{i \in S, j \in T, (i,j) \in \mathcal{E}^{st}} c(i,j) \tag{19}$$

An MRF that has such an s-t graph is called *graph-representable*[13] and can be solved in polynomial time using graph cuts [102]. 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 [48] 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 [48] and Goldberg-Tarjan style push-relabel algorithms [63]. Note that the min-cut problem and the max-flow problem are actually dual LP problems of each other [186]. Unfortunately, not all the MRFs are graph-representable. Previous works have been done to explore the class of graph-representable MRFs (*e.g.*, [21, 73, 102, 161]). 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

[11] Note that generations such as *multi-way cut* problem [35] which involves more than two terminal nodes are NP-hard.

[12] 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.

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

MRF is submodular. However, in numerous vision problems, more challenging energy functions are often required that do not satisfy the submodular condition (Eq. 23). The minimization of such non-submodular energy functions are NP-hard in general [28, 102] and an approximation algorithm would be required to approach the global optimum.

In vision community, more than two decades ago, a pioneer work [64] 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, such techniques did not draw much attention in the following decade since then, probably due to the fact that the model considered in [64] 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 [157, 26, 76], which proposed to use min-cut/max-flow techniques to minimize multi-label MRFs. 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 exist two main methodologies for solving multi-label MRFs based on graph cuts. The first methodology (referred to as *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 [73] to deal with arbitrary *convex* pairwise MRFs, which was further extended to submodular pairwise MRFs in [161]. The main idea of the second methodology (referred to as *move-making*) is to iteratively 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. Such a methodology provides an efficient approximate MRF optimization algorithm and the performance mainly depends on the size (denoted by $\mathcal{M}$) of the set of proposals at each iteration. ICM [13] 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. Based on graph cuts, $\mathcal{M}$ can be largely increased by considering the combination of two possible values for all the nodes ($\mathcal{M} = 2^{|\mathcal{V}|}$). A representative work on such a methodology is [27, 28], where *α-expansion* and *$\alpha\beta$-swap* were introduced to generalize binary graph cuts to handle pairwise MRFs with *metric* and/or *semi-metric* energy with optimum quality guarantee (*i.e.*, the ratio between the obtain energy and the global optimal energy is bounded by a factor).

Towards Non-submodular Functions

Graph cuts techniques has also been extended to deal with non-submodular binary energy functions. *Roof duality* was proposed in [66], which provides an LP relaxation (binary case of Eq. 22) approach to achieve 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). Such a method was efficiently implemented in [19], 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 [99]. 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*. 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. Later, two different techniques were introduced in order to extend QPBO towards achieving a complete solution. One is *probing* (called *QPBO-P*) [20, 156], 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*) [156], which starts from a complete 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 $\mathbf{y}$.

The QPBO techniques have been further combined with the *label-reduction* and *move-making* techniques presented previously to deal with multi-label MRFs. In [93], a multi-label MRF is converted into an equivalent binary MRF [73] and then QPBO techniques are employed to solve the linear relaxation of the obtained binary MRF. An elegant combination of QPBO and move-making techniques was proposed in [125], which is referred to as *fusion moves*. Different from previous move-making techniques such as α-expansion and $\alpha\beta$-swap, *fusion moves* fuses two arbitrary proposals of the full labeling by using QPBO and achieves a new labeling that is guaranteed to have an energy less or equal than the energies of both proposals.

Towards Improving Efficiency

We should also note that several methods have been proposed to increase the efficiency of graph-cuts-based MRF optimization. For example, a dynamic max-flow algorithm was proposed in [94, 95] to accelerate graph cuts when dealing with dynamics MRFs (*i.e.*, the potential functions vary over time, whereas the change between two successive instants is usually quite small), where the key idea is to reuse the flow obtained by solving the previous MRF so as to significantly reduce the computational time of min-cut. Another dynamic algorithm was also proposed in [80] 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 [109, 110], a primal-dual scheme based on linear programming relaxation was proposed for optimizing the MRF energy. This method can be viewed as a generalization of α-expansion and achieves a substantial speedup with respect to previous methods such as [28] and [108]. Two similar but simpler techniques with respect to that of [109, 110] were proposed in [2] to achieve a similar computational efficiency. Besides, an efficient algorithm based on max-flow and elimination techniques was introduced in [29] for the optimization of 4-neighborhood grid-like MRFs. Based on the primal-dual interpretation of the expansion algorithm introduced by [109, 110], [9] proposed an approach to optimize the choice of the move space (*i.e.*, the value of α for the expansion) for each iteration by exploiting the primal-dual gap, which was demonstrated experimentally to increase further the optimization efficiency.

4.2 Belief Propagation Algorithms

Belief propagation algorithms use local message passing to perform inference on graphical models. These methods provide an exact inference algorithm for tree-structured discrete graphical models, while an approximate solution can be achieved when a loopy graph is considered. For those loopy graphs with low tree-widths (Eq. 20) such as cycles, extended belief propagation methods such as *junction tree algorithm* [36, 1, 79] provide an efficient algorithm to perform exact inference. All these belief propagation algorithms have been adopted to perform MAP inference in MRF models for a variety of vision problems (*e.g.*, [42, 51, 175, 133, 8]).

4.2.1 Belief Propagation in Tree

Belief propagation (BP) [139, 206, 16] 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. These methods can be viewed as a special case of *dynamic programming* in graphical models [10, 34, 45].

The *min-sum* algorithm[14] is described in Algorithm 1 (B), where the factor graph representation [114, 16] is used, since as we mentioned in section 2.4, the factor graph makes the BP algorithm applicable to more cases compared to the classic min-sum algorithm applied on a usual pairwise MRF [51]. A representative vision model that can be efficiently solved by such a method is the pictorial model [46, 42] (see section 3.1.2). Note that *reparameterization* (also known as *equivalent transformation*) of the MRF energy (*e.g.*, [192, 98]) provides an alternative interpretation of belief propagation and leads to a memory-efficient implementation [98].

4.2.2 Loopy Belief Propagation

The tree-structured constraint limits the use of the standard belief propagation algorithm presented above and 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.*, [53, 51, 198, 43]) despite of the existence of loops. We refer the reader to [51, 198] 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. [179] showed empirically that the sequential scheme was significantly faster than the parallel one, while the performance of both methods was almost the same.

Substantial investment was made towards improving the efficiency of message passing by exploiting different types of structure regarding the graph and/or the potential functions. For example, an efficient method was proposed in [137] to reduce computational and memory cost for *robust truncated models* where a pairwise potential is equal to a constant for most of the state combination of the two nodes. [43] introduced a strategy for speeding up belief propagation for cases where the pairwise potential function only depends on the difference of the variables such as those defined in Eq. 15, an approach to accelerating the message passing in bipartite graphs (including grid-like MRFs in Fig. 3), and a multi-scale belief propagation scheme to perform inference in grid-like MRFs. Two speed-up techniques specifically for grid-like MRF models were also proposed in [141].

Despite the fact that LBP performed well for a number of vision applications such as [51, 175], 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.*, [193, 98, 106]) that will be presented in section 4.3 [178].

[14]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 [114, 16, 79] for more details.

4.2.3 Junction Tree Algorithm

Junction tree algorithm (JTA) is an exact inference method in arbitrary graphical models [36, 1, 79]. 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 this sense, we regard JTA as an extension of the standard belief propagation.

A junction tree corresponding to an MRF can be obtained by first triangulating the original graph and then find a maximal spanning tree for the maximal cliques contained in the triangulated graph. A short presentation of JTA is provided in C, from which we can easily notice that the complexity of the inference (*i.e.*, belief propagation) in a junction tree for a discrete MRF is exponential with respect to its *width* W. The width is defined as the maximum cardinal of the corresponding cliques over all nodes minus 1, *i.e.*:

$$W = \max_{i \in \mathcal{V}_J} |c_i| - 1 \tag{20}$$

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 [79]. 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 outer-planar graphs whose widths are equal to 2). For example, JTA was employed in [133] to deal with simultaneous localization and mapping (SLAM) problem, and was also adopted in [8] 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 [86] 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 (Eq. 10, 13), can be reformulated as the *integer linear programming (ILP)* [194] as follows:

$$\begin{aligned} \min_{\boldsymbol{\tau}} \quad & E(\boldsymbol{\theta}, \boldsymbol{\tau}) = \langle \boldsymbol{\theta}, \boldsymbol{\tau} \rangle = \sum_{i \in \mathcal{V}} \sum_{a \in \mathcal{X}_i} \theta_{i;a} \tau_{i;a} + \sum_{(i,j) \in \mathcal{E}} \sum_{(a,b) \in \mathcal{X}_i \times \mathcal{X}_j} \theta_{ij;ab} \tau_{ij;ab} \\ s.t. \quad & \boldsymbol{\tau} \in \boldsymbol{\tau}^{\mathcal{G}} = \left\{ \boldsymbol{\tau} \;\middle|\; \begin{array}{ll} \sum_{a \in \mathcal{X}_i} \tau_{i;a} = 1 & \forall\, i \in \mathcal{V} \\ \sum_{a \in \mathcal{X}_i} \tau_{ij;ab} = \tau_{j;b} & \forall\, \{i,j\} \in \mathcal{E}, b \in \mathcal{X}_j \\ \tau_{i;a} \in \{0,1\} & \forall\, i \in \mathcal{V}, a \in \mathcal{X}_i \\ \tau_{ij;ab} \in \{0,1\} & \forall\, \{i,j\} \in \mathcal{E}, (a,b) \in \mathcal{X}_i \times \mathcal{X}_j \end{array} \right\}. \end{aligned} \tag{21}$$

where $\theta_{i;a} = \theta_i(a)$, $\theta_{ij;ab} = \theta_{ij}(a,b)$, binary variables[15] $\tau_{i;a} = [x_u = a]$ 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 $\boldsymbol{\tau}$.

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

Unfortunately the above ILP problem is NP-hard in general. Numerous approximation algorithms of MRF optimization have been developed based on *Linear Programming (LP)* relaxation of such a problem in Eq. 21, 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 replacing the integer constraints in Eq. 21 by the non-negative constraints, *i.e.*:

$$\begin{aligned} \min_{\boldsymbol{\tau}} \quad & E(\boldsymbol{\theta}, \boldsymbol{\tau}) = \langle \boldsymbol{\theta}, \boldsymbol{\tau} \rangle = \sum_{i \in \mathcal{V}} \sum_{a \in \mathcal{X}_i} \theta_{i;a} \tau_{i;a} + \sum_{(i,j) \in \mathcal{E}} \sum_{(a,b) \in \mathcal{X}_i \times \mathcal{X}_j} \theta_{ij;ab} \tau_{ij;ab} \\ s.t. \quad & \boldsymbol{\tau} \in \hat{\boldsymbol{\tau}}^{\mathcal{G}} = \left\{ \boldsymbol{\tau} \;\middle|\; \begin{array}{ll} \sum_{a \in \mathcal{X}_i} \tau_{i;a} = 1 & \forall\, i \in \mathcal{V} \\ \sum_{a \in \mathcal{X}_i} \tau_{ij;ab} = \tau_{j;b} & \forall\, \{i,j\} \in \mathcal{E}, b \in \mathcal{X}_j \\ \tau_{i;a} \geq 0 & \forall\, i \in \mathcal{V}, a \in \mathcal{X}_i \\ \tau_{ij;ab} \geq 0 & \forall\, \{i,j\} \in \mathcal{E}, (a,b) \in \mathcal{X}_i \times \mathcal{X}_j \end{array} \right\}. \end{aligned} \tag{22}$$

For purposes of clarity, from now on, the term *MRF-MAP* will be used for the original MAP inference problem (Eq. 21) and *MRF-LP* for the relaxed one (Eq. 22).

It is generally infeasible to directly apply generic LP algorithms such as *interior point methods* [22] to solve MRF-LP problems corresponding to MRF models in computer vision [205], due to the fact that the number of variables involved in $\boldsymbol{\tau}$ is usually huge. Instead, many methods in the literature have been designed based on solving some *dual* to the MRF-LP problem in Eq. 22, *i.e.*, maximizing the lower bound of $E(\boldsymbol{\theta}, \boldsymbol{\tau})$ provided by the dual. One can cite for example the *min-sum diffusion* [112] and *augmenting DAG* [111] algorithms that were reviewed in [199], the *message passing* algorithm based on *block coordinate descent* proposed in [59], *tree-reweighted Message Passing (TRW)* techniques [193, 98] and *dual decomposition (MRF-DD)* [109, 107]. The tightening of the LP-relaxation has also been investigated towards achieving a better optimum of the MRF-MAP problem (*e.g.*, [169, 104, 136, 201]). Here, we review briefly the TRW and MRF-DD techniques as representatives.

4.3.1 Tree-reweighted Message Passing

Tree-reweighted max-product message passing (TRW) algorithms [193, 98] are well-explored MRF optimization methods. The key idea of TRW algorithms is to solve the MRF-LP problem via a dual problem based on convex combination of trees. Actually, the optimal values of such a dual problem and of the MRF-LP problem coincide, since strong duality holds [193]. Furthermore, in TRW algorithms, the LP relaxation (Eq. 22) is tight if a fix point of TRW algorithms satisfies a condition referred to as (strong) *tree agreement (TA)* [193], where a global optimal solution to the original MRF problem is achieved.

In [193], such an methodology was introduced to solve the MRF-MAP problem by using two different (edge-based and tree-based) message passing schemes, called *TRW-E* and *TRW-T*, respectively. These variants can be viewed as combinations of reparameterization and averaging operations on the MRF energy. However, both of the schemes do not guarantee the convergence of the algorithms and the value of the lower bound may fall into a loop. A sequential message passing scheme was proposed in [98], which is known as *TRW-S*. Different from TRW-E and TRW-T, the TRW-S algorithm updates messages in a sequential order instead of a parallel order. Such a difference introduce to the algorithm better convergence properties, *i.e.*, the lower bound will not decrease. TRW-S will attain a point that satisfies a condition referred to as *weak tree agreement (WTA)* [100] and the lower bound will not change any more since then[16].

[16] [98] observed in the experiments that TRW-S would finally converge to a fixed point but such a convergence

Although the global optimum of the dual problem satisfies WTA condition, the converse is not necessarily true and therefore TRW-S cannot guarantee the global maximum of the lower bound in general. Nevertheless, as demonstrated in [100], a WTA fixed point for the cases of binary pairwise MRFs always corresponds to the global maximum of the dual problem, and thus also corresponds to the global optimum of the MRF-LP problem. Furthermore, if a binary pairwise MRF is submodular, a WTA fixed point always achieves the global optimum of the MRF-MAP problem.

4.3.2 Dual Decomposition

In [106, 107], *dual-decomposition* [11] principle was introduced into the MRF optimization problem. The outcome was a general and powerful framework to minimize the MRF energy, which will be called *MRF-DD* in the remaining part of the survey. The key idea of MRF-DD is: instead of minimizing directly the energy of the original problem (referred to as *master* problem) that is too complex to solve directly, we decompose the master problem into a set of subproblems (referred to as *slave* problems). The main characteristic of these subproblems is that each of them is easier to solve both in terms of cardinality as well as in terms of convexity. Once such decomposition is achieved, the solution of the master problem is obtained by combining the solutions of the slaves problems. Such an idea can be summarized mathematically as following: based on a Lagrangian dual of the *MRF-MAP* problem in Eq. 21, the sum of the minima of the slave problems that are obtained by the decomposition of the master problem 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 master problem can be extracted from the Lagrangian solutions[17].

Such a MRF optimization framework possesses a great flexibility, generality and convergence property:

1. The Lagrangian dual problem can be globally optimized due to the convexity of the dual function. The solution obtained by the MRF-DD algorithm satisfies *weak tree agreement (WTA)* condition[18], while a solution satisfying WTA condition is not necessarily the optimum to the Lagrangian dual. The properties of *tree agreement* and *weak tree agreement* fix points [100] are also applicable within the MRF-DD method.

2. Different decompositions of the master problem can be considered to deal with MRF-MAP problem. Each of such decompositions leads to a certain relaxation of the MRF-MAP problem. Interestingly, when the master problem is decomposed into a set of trees, the Lagrangian relaxation employed by MRF-DD is equivalent to the LP relaxation in Eq. 22, which is exactly the problem TRW algorithms aim to solve[19]. However, within MRF-DD framework, one can consider more sophisticated decompositions to tighten the relaxation (*e.g.*, decompositions based on outer-planar graphs [8] and K-fan graphs [83]). To this end, a very useful theoretical conclusion has been drawn in [107] which provides an approach to comparing the tightness between two different decompositions.

3. Only MAP inference in slave problems are required and there is no constraints on how such an inference is done. As a result, one can apply specific optimization algorithms to

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.

[17] [107] provides a detailed discussion on different approaches to obtaining a feasible solution of the master problem from the solution of the slave problems after solving the Lagrangian dual.

[18] WTA condition can be easily extended to the cases where one or more slave problems are not tree-structured.

[19] The main difference between MRF-DD and TRW algorithms consists in the mechanism of the update of dual variables. The former relies on the optimal solution of slave problems while the latter is based on the min-marginals of the trees corresponding to slave problems.

solve slave problems and even different optimization algorithms for different slave problems. The natural outcome of such a property is high flexibility for designing new graph-based optimization algorithms based on such a dual decomposition framework. A number of elegant applications have been proposed in the literature, which include the graph matching method proposed in [184], the higher-order MRF inference method developed in [105], and the algorithm for joint segmentation and appearance histogram models optimization introduced in [190].

However, computational cost is the main drawback of the MRF-DD algorithm. Reducing the running time for the convergence is an open problem and there are various techniques that have been proposed in the literature. For example, two approaches were proposed in [103] to speed-up LP-based algorithms. 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. [78] 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 [171] to solve the slave problems in parallel.

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 [151, 153], 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.*, [90, 148, 132, 37, 118]), various techniques also have been proposed to deal with more general MRF models (*e.g.*, [122, 143, 105, 74, 47]). 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 [150] to perform order reduction for models of any order, by introducing auxiliary variables to substitute products of variables[20]. 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 [21, 3], since our final interest is solving higher-order models. In [3], 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 [102], in order to construct s-t graph for a third-order binary MRF.

[20] Here, we consider binary higher-order MRFs and their energy functions can be represented in form of *pseudo-Boolean functions* [21].

This reduction method was studied from an algebraic viewpoint in [49] and led to some interesting conclusions towards extending this method to models of an arbitrary order. Based on these works, [74, 75] proposed a generalized technique that can reduce any higher-order binary MRF into a pairwise one, which can then be solved by QBPO. Furthermore, [74, 75] also extended such a technique to deal with multi-label MRFs by using fusion moves [125]. 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), [55] proposed to approach order reduction as a optimization problem, where different factors are allowed to choose different reduction methods towards optimizing an objective function defined using a special graph (referred to as *order reduction inference graph*). In the same line of research, [47] proposed to perform order reduction on a group of higher-order terms at the same time instead of on each terms independently [74, 75], which has been demonstrated both theoretically and experimentally to lead to better performance compared to [74, 75].

Graph-cuts techniques have also been considered to cope either with specific vision problems or certain classes of higher-order models. For example, [90, 91] characterized a class of higher-order potentials (*i.e.*, $\mathcal{P}^n$ Potts model) for which the optimal expansion and swap moves can be computed efficiently in polynomial time, and proposed an efficient graph-cuts-based method for solving such models. Such a technique was further extended in [87, 88] to a wider class of higher-order models (*i.e.*, robust $\mathcal{P}^n$ model). Graph-cuts-based approaches were also proposed [118, 119] and in [118, 119] 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, they better capture the characteristics of the problems and are able to solve the problems relatively efficiently (*e.g.*, they often cannot be solved by a general inference methods).

4.4.2 Belief-propagation-based Methods

As we 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 employed in [122] 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, the use of *distance transform* techniques [18, 43] significantly improves the efficiency of the message passing process in [122]. [142, 143] and [180] 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 [129] by exploiting the fact that a clique potential often consists of a sum of potentials each of which involves only a sub-clique of variables. The expected time of the message passing was further reduced in [44].

4.4.3 Dual Methods

The LP relaxation formulation in Eq. 22 can be generalized to the cases of higher-order MRFs. Such a generalization was studied in [200, 201], where *min-sum diffusion* [112] was adopted to achieve a method for optimizing the energy of higher-order MRFs, which is referred to as *n-ary*

min-sum diffusion[21]. Recently, such techniques were adopted in [208] to efficiently solve in a parallel/distributed fashion higher-order MRF models of triangulated planar structure.

The *Dual-decomposition* framework [11, 109], which has been presented in section 4.3.2, can also be adopted to deal with higher-order MRFs. This was first demonstrated in [105], where inference algorithms were introduced for solving a wide class of higher-order potential referred to as *pattern-based potentials*[22]. Also based on the dual-decomposition framework, [196] proposed to solved their higher-order MRF model by decomposing the original problem into a series of subproblems each of which corresponds to a factor tree. In [207], such a framework was combined with order-reduction [74, 75] and QPBO techniques [99] to solve higher-order graph-matching problems.

Exploitation of the Sparsity of Potentials

Lastly, we note that the exploitation of the sparsity of potentials is explicitly or implicitly employed in many of the above higher-order inference methods. In this direction, [154] proposed a compact representation for "sparse" higher-order potentials (except a very small subset, the labelings are almost impossible so as to have the same high energy) to convert a higher-order model into a pairwise one so that pairwise MRF inference methods such as graph cuts can be employed to solve the problem. Due to the "sparseness", only a small number of auxiliary variables are required for the order reduction process. In the same line of research, [89] studied and characterized some families of higher-order potentials (*e.g.*, $\mathcal{P}^n$ Potts model [91]) that can be represented compactly as upper or lower envelopes of linear functions. Furthermore, it was demonstrated that these higher-order models can be converted into pairwise models with the addition of a small number of auxiliary variables.

5 Conclusion

In order to conclude this survey, let us first recall to the reader 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 wider vision problems of low, mid and high-level. On the other side, 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 suggests that MRFs will keep being a major research topic and offer more promise than ever before.

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 [161]. 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 $\forall\, x_j^1, x_j^2 \in \mathcal{X}_j \ s.t.\ x_j^1 \leq x_j^2$,

$$\theta_{ij}(x_i^1, x_j^1) + \theta_{ij}(x_i^2, x_j^2) \leq \theta_{ij}(x_i^1, x_j^2) + \theta_{ij}(x_i^2, x_j^1), \tag{23}$$

[21] The method was originally called *n-ary max-sum diffusion* in [200, 201] due to the fact that a maximization of objective function was considered.

[22] For example, $\mathcal{P}^n$ Potts model [91] is a sub-class of *pattern-based potentials*

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

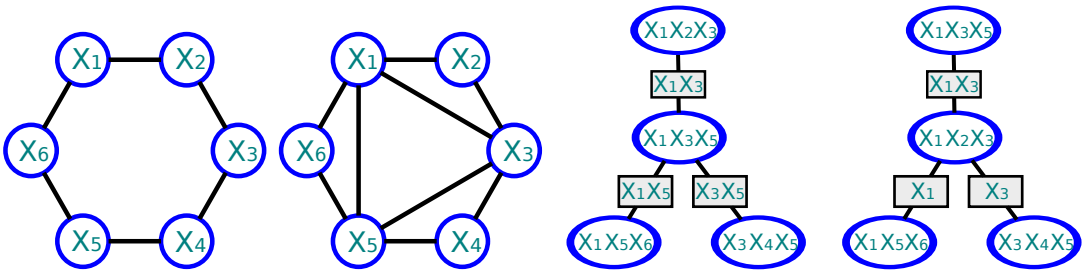
$$\theta_{ij}(0,0) + \theta_{ij}(1,1) \leq \theta_{ij}(0,1) + \theta_{ij}(1,0) \tag{24}$$

One can refer to [102] for generalizing the submodularity to higher-order MRFs.

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 $\mathbf{x}^{\text{opt}} = \arg\min_{\mathbf{x}} \sum_{f \in \mathcal{F}} \theta_f(x_f)$
  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 $\boldsymbol{\mathcal{C}}$ *s.t.* $\boldsymbol{\mathcal{C}}(i)$ 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 $\mathcal{C} \leftarrow \boldsymbol{\mathcal{C}}(i)$
    if $i \in \mathcal{V}$ **then**
      if $|\mathcal{C}| > 0$ **then**
        $m_{i \rightarrow p}(x_i) \leftarrow \sum_{j \in \mathcal{C}} m_{j \rightarrow i}(x_i)$
      else
        $m_{i \rightarrow p}(x_i) \leftarrow 0$
      end if
    else
      if $|\mathcal{C}| > 0$ **then**
        $m_{i \rightarrow p}(x_p) \leftarrow \min_{x_{\mathcal{C}}}(\phi(x_i) + \sum_{j \in \mathcal{C}} m_{j \rightarrow i}(x_j))$
        $s_i(x_p) \leftarrow \arg\min_{x_{\mathcal{C}}}(\phi(x_i) + \sum_{j \in \mathcal{C}} m_{j \rightarrow i}(x_j))$
      else
        $m_{i \rightarrow p}(x_p) \leftarrow \phi(x_p)$ {p is the unique variable contained in factor i in this case.}
      end if
    end if
  end for
  $x_{\hat{r}}^{\text{opt}} \leftarrow \arg\min_{x_{\hat{r}}} \sum_{j \in \boldsymbol{\mathcal{C}}(\hat{r})} m_{j \rightarrow \hat{r}}(x_{\hat{r}})$
  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 $\mathcal{C} \leftarrow \boldsymbol{\mathcal{C}}(i)$
      $x_{\mathcal{C}}^{\text{opt}} \leftarrow s_i(x_p)$
    end if
  end for
  return $\mathbf{x}^{\text{opt}}$



(a) (b) (c) (d)

Figure 6: 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.

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: $\mathcal{P}_{\text{send}} = \text{NodeOrdering}(\mathcal{T}, \hat{r})$, 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
      $\mathcal{P}_{\text{send}} \leftarrow (\text{NodeOrdering}(\mathcal{T}, \hat{r}), \mathcal{P}_{\text{send}})$ {$\mathcal{P}_{\text{send}}$ is ordered from left to right}
    end for
  end if
  return $\mathcal{P}_{\text{send}}$

C Junction Tree Algorithm

Let us introduce some necessary notions and properties about junction trees and then discuss briefly the corresponding inference algorithm. For a clique set $\mathcal{C}$, the corresponding *clique tree* is defined as a tree-structured graph $\mathcal{G}_J$ with node set $\mathcal{V}_J$ and edge set $\mathcal{E}_J$ where each node i ($i \in \mathcal{V}_J$) represents a clique $c_i \in \mathcal{C}$. A *junction tree* is a clique tree which processes the junction tree property: for every pair of cliques c_i and c_j in $\mathcal{G}_J$, $c_i \cap c_j$ is contained in all the cliques on the (unique) path between c_i and c_j. The junction tree property ensures that local consistency implies global consistency so that local message passing process can produce exact inference. The example in Fig. 6 provides two clique trees (Fig. 6(c) and (d)) corresponding to the undirected graph in Fig. 6(b), where we use square boxes to explicitly represent the separators each of which is associated to an edge and denotes the intersection of the two cliques connected by the edge. We can easily verify that the clique tree in Fig. 6(c) is a *junction tree*, while the other one in Fig. 6(d) is not.

There are two important properties about junction trees [79], which are useful for the construction of a junction tree given an undirected graphical model:

1. An undirected graph has a *junction tree* if and only if it is triangulated (*i.e.*, there is no

chordless[23] cycle in the graph.

2. A clique tree is a junction tree if and only if it is a maximal spanning tree which is a clique tree that has the maximal weight (*i.e.*, $\sum_{i,j\in\mathcal{E}_J} |c_i \cap c_j|$) over all possible trees connecting the considered cliques.

Hence, for a given undirected graph (*e.g.*, Fig. 6(a)), we can first triangulate[24] it (*e.g.*, Fig. 6(b)), and then find a maximal spanning tree to form a junction tree for the maximal cliques contained in this triangulated graph. This operation will produce a junction tree for the undirected graph (*e.g.*, Fig. 6(c)). For each clique c in the original graph, the associated clique potential θ_c is accumulated to the potential $\hat{\theta}_i$ of one and only one node i in the junction tree such that c is included in the clique c_i corresponding to node i (*i.e.*, $c \subseteq c_i$).

Without considering optimality of the generated junction tree[25], the triangulation can be done easily using *undirected graph elimination* algorithm [79]. This method successively eliminates the nodes in a graph by connecting the remaining neighbors of the node and removing the node as well as the edge connected to it from the graph. The second step, *i.e.*, the finding of a maximal spanning tree, can be easily performed using greedy algorithms such as Kruskal's algorithm [34].

The energy[26] of a junction tree is defined as a sum of the potentials of the cliques corresponding to the nodes:

$$E(\mathbf{x}) = \sum_{i\in\mathcal{V}_J} \hat{\theta}_i(x_{c_i}) \tag{25}$$

where c_i denotes the clique corresponding to node i of the junction tree. Due to the junction tree property, we can perform local message passing in the junction tree to do the inference, which is similar to standard belief propagation in factor trees. Interestingly, nodes in junction trees can be regarded as factor nodes in factor trees, while separators in junction trees can be regarded as usual nodes (may corresponding to a set of variables) in factor trees. Then the belief propagation scheme in the junction tree can be obtained easily from the one for the factor tree (see Algorithm 1). Hence, we do not present the message passing process here to avoid redundancy and refer the reader to [1, 79] for details.

References

[1] S.M. Aji and R.J. McEliece. The generalized distributive law. *IEEE Transactions on Information Theory*, 46(2):325–343, March 2000.

[2] Karteek Alahari, Pushmeet Kohli, and Philip H. S. Torr. Reduce, reuse & recycle: Efficiently solving multi-label MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[3] Asem M. Ali, Aly A. Farag, and Georgy L. Gimel'farb. Optimizing binary MRFs with higher order cliques. In *European Conference on Computer Vision (ECCV)*, 2008.

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

[24] For directed graphical models, a *moralization* process [79] is to be applied prior to the triangulation in order to transform the directed graph to an undirected graph.

[25] Note that there may exist several such junction trees corresponding to an undirected graph. As we will discuss below, the optimality of a junction tree is related to its *width*. However, it is generally an NP-hard problem to find an optimal junction tree [79].

[26] The joint probability of a junction tree is defined as a product of potential functions corresponding to the nodes, which is similar to that of a factor graph in Eq. 11 and is not presented here for the purpose of compactness.

[4] Amir A. Amini, Terry E. Weymouth, and Ramesh C. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 12(9):855–867, 1990.

[5] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. Pictorial structures revisited: People detection and articulated pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[6] Dragomir Anguelov, Praveen Srinivasan, Hoi-Cheung Pang, Daphne Koller, Sebastian Thrun, and James Davis. The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. In *Advances in Neural Information Processing Systems (NIPS)*, 2004.

[7] A. Ayvaci and S. Soatto. Motion segmentation with occlusions on the superpixel graph. In *ICCV Workshop on Dynamical Vision*, October 2009.

[8] Dhruv Batra, A. C. Gallagher, Devi Parikh, and Tsuhan Chen. Beyond Trees: MRF Inference via Outer-Planar Decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[9] Dhruv Batra and Pushmeet Kohli. Making the right moves: Guiding alpha-expansion using local primal-dual gaps. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

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

[11] Dimitri P. Bertsekas. *Nonlinear Programming (Second Edition).* Athena Scientific, 1999.

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

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

[14] A. Besbes, N. Komodakis, G. Langs, and N. Paragios. Shape priors and discrete mrfs for knowledge-based segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[15] Christopher M. Bishop. *Neural networks for pattern recognition.* Oxford university press, 1995.

[16] Christopher M. Bishop. *Pattern recognition and machine learning (Information Science and Statistics).* Springer, 2006.

[17] Andrew Blake and Andrew Zisserman. *Visual Reconstruction.* MIT Press, 1987.

[18] Gunilla Borgefors. Distance transformations in digital images. *Computer vision, graphics, and image processing*, 34(3):344–371, 1986.

[19] Endre Boros, P. L. Hammer, and X. Sun. Network flows and minimization of quadratic pseudo-Boolean functions. *RUTCOR Research Report RRR 17-1991*, 1991.

[20] Endre Boros, P. L. Hammer, and Gabriel Tavares. Preprocessing of unconstrained quadratic binary optimization. *RUTCOR Research Report RRR 10-2006*, 2006.

[21] Endre Boros and Peter L. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123:155–225, 2002.

[22] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[23] Yuri Boykov and Gareth Funka-Lea. Graph Cuts and Efficient N-D Image Segmentation. *International Journal of Computer Vision (IJCV)*, 70(2):109–131, November 2006.

[24] Yuri Boykov and Marie-Pierre Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *IEEE International Conference on Computer Vision (ICCV)*, 2001.

[25] Yuri Boykov and Vladimir Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *IEEE International Conference on Computer Vision (ICCV)*, 2003.

[26] Yuri Boykov, Olga Veksler, and Ramin Zabih. Markov random fields with efficient approximations. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1998.

[27] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. In *International Conference on Computer Vision (ICCV)*, 1999.

[28] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 23(11):1222–1239, 2001.

[29] Peter Carr and Richard Hartley. Minimizing energy functions on 4-connected lattices using elimination. In *IEEE International Conference on Computer Vision (ICCV)*, 2009.

[30] Antonin Chambolle. Total variation minimization and a class of binary mrf models. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, 2005.

[31] Tony F. Chan and Jianhong Shen. *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. Society for Industrial and Applied Mathematics (SIAM), 2005.

[32] P. B. Chou, P. R. Cooper, M. J. Swain, C. M. Brown, and L. E. Wixson. Probabilistic network inference for cooperative high and low level vision. In R. Chellappa and A. Jain, editors, *Markov Random Fields: Theory and Applications*, pages 211–243. Academic Press, 1993.

[33] Paul B. Chou and Christopher M. Brown. The theory and practice of Bayesian image labeling. *International Journal of Computer Vision (IJCV)*, 4(3):185–210, 1990.

[34] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms (Third Edition)*. The MIT press, 2009.

[35] E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis. The complexity of multiway cuts (extended abstract). In *ACM Symposium on Theory of Computing (STOC)*, 1992.

[36] A. P. Dawid. Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing*, 2(1):25–36, 1992.

[37] Andrew Delong, Anton Osokin, Hossam N. Isack, and Yuri Boykov. Fast approximate energy minimization with label costs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[38] M. Egmont-Petersen, D. de Ridderb, and H. Handelsc. Image processing with neural networks-a review. *Pattern Recognition*, 35(10):2279–2301, October 2002.

[39] Marcin Eichner and Vittorio Ferrari. Better appearance models for pictorial structures. In *British Machine Vision Conference (BMVC)*, 2009.

[40] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, Dordrecht, 1996.

[41] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(9):1627–1645, September 2010.

[42] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Pictorial Structures for Object Recognition. *International Journal of Computer Vision (IJCV)*, 61(1):55–79, January 2005.

[43] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient Belief Propagation for Early Vision. *International Journal of Computer Vision (IJCV)*, 70(1):41–54, May 2006.

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

[45] Pedro F. Felzenszwalb and Ramin Zabih. Dynamic programming and graph algorithms in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(4):721–740, April 2011.

[46] M.A. Fischler and R.A. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22(1):67–92, 1973.

[47] Alexander Fix, Aritanan Gruber, Endre Boros, and Ramin Zabih. A Graph Cut Algorithm for Higher-order Markov Random Fields. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[48] L. R. Ford and D. R. Fulkerson. *Flows in networks*. Princeton University Press, 1962.

[49] Daniel Freedman and Petros Drineas. Energy Minimization via Graph Cuts: Settling What is Possible. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[50] William T. Freeman, Thouis R. Jones, and Egon C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.

[51] William T. Freeman, Egon C. Pasztor, and Owen T. Carmichael. Learning low-level vision. *International Journal of Computer Vision (IJCV)*, 40(1):25–47, 2000.

[52] Brendan J. Frey. *Graphical models for machine learning and digital communication*. MIT Press, 1998.

[53] Brendan J. Frey and David J. C. MacKay. A revolution: Belief propagation in graphs with cycles. In *Advances in Neural Information Processing Systems (NIPS)*, 1998.

[54] B. Fulkerson, A. Vedaldi, and S. Soatto. Class segmentation and object localization with superpixel neighborhoods. In *International Conference on Computer Vision (ICCV)*, 2009.

[55] Andrew C. Gallagher, Dhruv Batra, and Devi Parikh. Inference for Order Reduction in Markov Random Fields. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[56] Dan Geiger, Thomas Verma, and Judea Pearl. Identifying independence in bayesian networks. *Networks*, 20:507–534, 1990.

[57] Arthur Gelb, editor. *Applied optimal estimation*. MIT Press, 1974.

[58] Stuart Geman and Donald Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 6(6):721–741, 1984.

[59] Amir Globerson and Tommi Jaakkola. Fixing Max-Product: Convergent Message Passing Algorithms for MAP LP-Relaxations. In *Advances in Neural Information Processing Systems (NIPS)*, 2007.

[60] Ben Glocker, T. Hauke Heibel, Nassir Navab, Pushmeet Kohli, and Carsten Rother. TriangleFlow: Optical Flow with Triangulation-Based Higher-Order Likelihoods. In *European Conference on Computer Vision (ECCV)*, 2010.

[61] Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab, and Nikos Paragios. Dense image registration through MRFs and efficient linear programming. *Medical Image Analysis*, 12(6):731–741, December 2008.

[62] Ben Glocker, Nikos Paragios, Nikos Komodakis, Georgios Tziritas, and Nassir Navab. Optical flow estimation with uncertainties through dynamic MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[63] A. V. Goldberg and R. E. Tarjan. A new approach to the maximum-flow problem. *Journal of the ACM (JACM)*, 35(4):921–940, 1988.

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

[65] Laura Gui, Jean-Philippe Thiran, and Nikos Paragios. Cooperative Object Segmentation and Behavior Inference in Image Sequences. *International Journal of Computer Vision (IJCV)*, 84(2):146–162, June 2008.

[66] P. L. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 0-1 optimization. *Mathematical Programming*, 28(2):121–155, 1984.

[67] J. M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. *unpublished*, 1971.

[68] Xuming He, Richard S. Zemel, and Miguel A. Carreira-Perpinan. Multiscale conditional random fields for image labeling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[69] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3D medical image segmentation: a review. *Medical Image Analysis*, 13(4):543–63, August 2009.

[70] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 15(12):1217–1232, 1993.

[71] A. Hervieu, P. Bouthemy, and J.-P. Le Cadre. A HMM-based method for recognizing dynamic video contents from trajectories. In *IEEE International Conference on Image Processing (ICIP)*, 2007.

[72] Michael Isard. PAMPAS: real-valued graphical models for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[73] Hiroshi Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 25(10):1333–1336, 2003.

[74] Hiroshi Ishikawa. Higher-order clique reduction in binary graph cut. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2009.

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

[76] Hiroshi Ishikawa and Davi Geiger. Segmentation by grouping junctions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1998.

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

[78] Vladimir Jojic, Stephen Gould, and Daphne Koller. Accelerated dual decomposition for MAP inference. In *International Conference on Machine Learning (ICML)*, 2010.

[79] Michael I. Jordan. *An introduction to probabilistic graphical models*. In preparation, 2007.

[80] Olivier Juan and Yuri Boykov. Active Graph Cuts. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[81] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME - Journal of Basic Engineering*, 82:35–45, 1960.

[82] Evangelos Kalogerakis, Aaron Hertzmann, and Karan Singh. Learning 3d mesh segmentation and labeling. *ACM Transactions on Graphics (TOG)*, 29(4):102:1–102:12, 2010.

[83] Jorg Hendrik Kappes, Stefan Schmidt, and Christoph Schnorr. MRF Inference by k-Fan Decomposition and Tight Lagrangian Relaxation. In *European Conference on Computer Vision (ECCV)*, 2010.

[84] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision (IJCV)*, 1(4):321–331, January 1988.

[85] Jaemin Kim and J. W. Woods. Spatio-temporal adaptive 3-D Kalman filter for video. *IEEE Transactions on Image Processing (TIP)*, 6(3):414–424, January 1997.

[86] Uffe Kjæ rulff. Inference in bayesian networks using nested junction trees. In *Proceedings of the NATO Advanced Study Institute on Learning in graphical models*, 1998.

[87] Pushmeet Kohli, Lubor Ladicky, and Philip H. S. Torr. Robust higher order potentials for enforcing label consistency. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[88] Pushmeet Kohli, L'ubor Ladický, and Philip H. S. Torr. Robust Higher Order Potentials for Enforcing Label Consistency. *International Journal of Computer Vision (IJCV)*, 82(3):302–324, January 2009.

[89] Pushmeet Kohli and M. Pawan Kumar. Energy minimization for linear envelope MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

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

[91] Pushmeet Kohli, M. Pawan Kumar, and Philip H. S. Torr. P3 & beyond: move making algorithms for solving higher order functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 31(9):1645–1656, September 2009.

[92] Pushmeet Kohli, Jonathan Rihan, Matthieu Bray, and Philip H. S. Torr. Simultaneous Segmentation and Pose Estimation of Humans Using Dynamic Graph Cuts. *International Journal of Computer Vision (IJCV)*, 79(3):285–298, January 2008.

[93] Pushmeet Kohli, Alexander Shekhovtsov, Carsten Rother, Vladimir Kolmogorov, and Philip H. S. Torr. On partial optimality in multi-label MRFs. In *International Conference on Machine Learning (ICML)*, 2008.

[94] Pushmeet Kohli and Philip H. S. Torr. Efficiently solving dynamic Markov random fields using graph cuts. *IEEE International Conference on Computer Vision (ICCV)*, 2005.

[95] Pushmeet Kohli and Philip H. S. Torr. Dynamic graph cuts for efficient inference in Markov Random Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(12):2079–2088, December 2007.

[96] Pushmeet Kohli and Philip H. S. Torr. Measuring uncertainty in graph cut solutions. *Computer Vision and Image Understanding (CVIU)*, 112(1):30–38, October 2008.

[97] Daphne Koller and Nir Friedman. *Probabilistic graphical models: Principles and techniques*. The MIT Press, 2009.

[98] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 28(10):1568–1583, October 2006.

[99] Vladimir Kolmogorov and Carsten Rother. Minimizing nonsubmodular functions with graph cuts - a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(7):1274–1279, July 2007.

[100] Vladimir Kolmogorov and Martin J. Wainwright. On the optimality of tree-reweighted max-product message-passing. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2005.

[101] Vladimir Kolmogorov and Ramin Zabih. Multi-camera Scene Reconstruction via Graph Cuts. In *European Conference on Computer Vision (ECCV)*, 2002.

[102] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 26(2):147–159, February 2004.

[103] Nikos Komodakis. Towards More Efficient and Effective LP-Based Algorithms for MRF Optimization. In *European Conference on Computer Vision (ECCV)*, 2010.

[104] Nikos Komodakis and Nikos Paragios. Beyond Loose LP-relaxations: Optimizing MRFs by Repairing Cycles. In *European Conference on Computer Vision (ECCV)*, 2008.

[105] Nikos Komodakis and Nikos Paragios. Beyond pairwise energies: Efficient optimization for higher-order MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[106] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. MRF Optimization via Dual Decomposition: Message-Passing Revisited. In *IEEE International Conference on Computer Vision (ICCV)*, 2007.

[107] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. MRF energy minimization and beyond via dual decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(3):531–552, March 2011.

[108] Nikos Komodakis and Georgios Tziritas. Approximate labeling via graph cuts based on linear programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(8):1436–1453, August 2007.

[109] Nikos Komodakis, Georgios Tziritas, and Nikos Paragios. Fast, Approximately Optimal Solutions for Single and Dynamic MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[110] Nikos Komodakis, Georgios Tziritas, and Nikos 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):14–29, October 2008.

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

[112] V. A. Kovalevsky and V. K. Koval. A diffusion algorithm for decreasing energy of max-sum labeling problem. Technical report, Glushkov Institute Of Cybernetics, Kiev, USSR, 1975.

[113] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in Neural Information Processing Systems (NIPS)*. 2011.

[114] Frank R. Kschischang, Brendan J. Frey, and Hans-Andrea Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.

[115] Sanjiv Kumar and Martial Hebert. Discriminative fields for modeling spatial dependencies in natural images. In *Advances in Neural Information Processing Systems (NIPS)*, 2004.

[116] Dongjin Kwon, Kyong Joon Lee, Il Dong Yun, and Sang Uk Lee. Nonrigid Image Registration Using Dynamic Higher-Order MRF Model. In *European Conference on Computer Vision (ECCV)*, 2008.

[117] Lubor Ladicky, Christopher Russell, Pushmeet Kohli, and Philip H. S. Torr. Associative hierarchical CRFs for object class image segmentation. In *IEEE International Conference on Computer Vision (ICCV)*, 2009.

[118] Lubor Ladicky, Christopher Russell, Pushmeet Kohli, and Philip H. S. Torr. Graph Cut based Inference with Co-occurrence Statistics. In *European Conference on Computer Vision (ECCV)*, 2010.

[119] Lubor Ladicky, Christopher Russell, Pushmeet Kohli, and Philip H. S. Torr. Inference Methods for CRFs with Co-occurrence Statistics. *International Journal of Computer Vision (IJCV)*, 2011.

[120] Lubor Ladicky, Paul Sturgess, Karteek Alahari, Chris Russell, and Philip H. S. Torr. What, where & how many? combining object detectors and crfs. *European Conference on Computer Vision (ECCV)*, 2010.

[121] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *International Conference on Machine Learning (ICML)*, 2001.

[122] Xiangyang Lan, Stefan Roth, Daniel P. Huttenlocher, and Michael J. Black. Efficient Belief Propagation with Learned Higher-Order Markov Random Fields. In *European Conference on Computer Vision (ECCV)*, 2006.

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

[124] D. Lee and T. Pavlidis. One-dimensional regularization with discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 10(6):822–829, 1988.

[125] Victor Lempitsky, Carsten Rother, Stefan Roth, and Andrew Blake. Fusion moves for markov random field optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(8):1392–1405, August 2010.

[126] Alex Levinshtein, Cristian Sminchisescu, and Sven Dickinson. Optimal Contour Closure by Superpixel Grouping. In *European Conference on Computer Vision (ECCV)*, 2010.

[127] Stan Z. Li. *Markov random field modeling in image analysis (Third Edition)*. Springer, 2009.

[128] Ce Liu, Jenny Yuen, and Antonio Torralba. SIFT Flow: Dense Correspondence across Scenes and Its Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(5):978–994, 2011.

[129] Julian J. Mcauley and Tiberio S. Caetano. Faster Algorithms for Max-Product Message-Passing. *Journal of Machine Learning Research*, 12:1349–1388, 2011.

[130] M. A. Moni and A. B. M. Shawkat Ali. HMM based hand gesture recognition: A review on techniques and approaches. In *IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, 2009.

[131] Alastair P. Moore, Simon J. D. Prince, and Jonathan Warrell. "Lattice Cut" - Constructing superpixels using layer constraints. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[132] Sebastian Nowozin and Christoph H. Lampert. Global connectivity potentials for random field models. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[133] Mark A. Paskin. Thin Junction Tree Filters for Simultaneous Localization and Mapping. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

[134] Vladimir Pavlovic. *Dynamic Bayesian Networks for Information Fusion with Application to Human-Computer Interfaces*. PhD thesis, University of Illinois at Urbana-Champaign, 1999.

[135] M. Pawan Kumar. *Combinatorial and Convex Optimization for Probabilistic Models in Computer Vision*. PhD thesis, Oxford Brookes University, 2008.

[136] M. Pawan Kumar, Vladimir Kolmogorov, and Philip H. S. Torr. An Analysis of Convex Relaxations for MAP Estimation of Discrete MRFs. *Journal of Machine Learning Research*, 10:71–106, 2009.

[137] M. Pawan Kumar and Philip H. S. Torr. Fast memory-efficient generalized belief propagation. *European Conference on Computer Vision (ECCV)*, 2006.

[138] M. Pawan Kumar, Philip H. S. Torr, and Andrew Zisserman. Learning Layered Pictorial Structures from Video. In *The Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, 2004.

[139] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[140] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, second edition, 2009.

[141] Kersten Petersen, Janis Fehr, and Hans Burkhardt. Fast generalized belief propagation for MAP estimation on 2D and 3D grid-like markov random fields. *DAGM-Symposium*, pages 41–50, 2008.

[142] Brian Potetz. Efficient Belief Propagation for Vision Using Linear Constraint Nodes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[143] Brian Potetz and Tai Sing Lee. Efficient belief propagation for higher-order cliques using linear constraint nodes. *Computer Vision and Image Understanding (CVIU)*, 112(1):39–54, October 2008.

[144] R. B. Potts. Some generalized order-disorder transitions. *Proceedings of the Cambridge Philosophical Society*, 48:106–109, 1952.

[145] Ariadna Quattoni, Michael Collins, and Trevor Darrell. Conditional random fields for object recognition. In *Advances in Neural Information Processing Systems (NIPS)*, 2004.

[146] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[147] Deepu Rajan and Subhasis Chaudhuri. An MRF-based approach to generation of super-resolution images from blurred observations. *Journal of Mathematical Imaging and Vision*, 16(1):5–15, 2002.

[148] Srikumar Ramalingam, Pushmeet Kohli, Karteek Alahari, and Philip H. S. Torr. Exact inference in multi-label CRFs with higher order cliques. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[149] J K Romberg, H Choi, and R G Baraniuk. Bayesian tree-structured image modeling using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing (TIP)*, 10(7):1056–1068, January 2001.

[150] I. G. Rosenberg. Reduction of bivalent maximization to the quadratic case. *Cahiers du Centre d'etudes de Recherche Operationnelle*, 17:71–74, 1975.

[151] Stefan Roth and Michael J. Black. Fields of Experts: A Framework for Learning Image Priors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[152] Stefan Roth and Michael J. Black. On the Spatial Statistics of Optical Flow. *International Journal of Computer Vision (IJCV)*, 74(1):33–50, January 2007.

[153] Stefan Roth and Michael J. Black. Fields of Experts. *International Journal of Computer Vision (IJCV)*, 82(2):205–229, January 2009.

[154] Carsten Rother, Pushmeet Kohli, Wei Feng, and Jiaya Jia. Minimizing sparse higher order energy functions of discrete variables. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[155] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts. *ACM Transactions on Graphics (TOG)*, 23(3):309–314, 2004.

[156] Carsten Rother, Vladimir Kolmogorov, Victor Lempitsky, and Martin Szummer. Optimizing Binary MRFs via Extended Roof Duality. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[157] Sebastien Roy and Ingemar J. Cox. A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem. In *IEEE International Conference on Computer Vision (ICCV)*, 1998.

[158] Sebastien Roy and Venu Govindu. Mrf solutions for probabilistic optical flow formulations. In *International Conference on Pattern Recognition (ICPR)*, 2000.

[159] Havard Rue and Leonhard Held. *Gaussian Markov Random Fields: Theory and Applications*. Chapman & HALL/CRC, 2005.

[160] Ruslan Salakhutdinov. Learning in Markov random fields using tempered transitions. In *Advances in Neural Information Processing Systems (NIPS)*, 2009.

[161] Dmitrij Schlesinger and Boris Flach. Transforming an arbitrary minsum problem into a binary one. Technical report, Dresden University of Technology, 2006.

[162] Dieter Seghers, Dirk Loeckx, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Minimal shape and intensity cost path segmentation. *IEEE Transactions on Medical Imaging (TMI)*, 26(8):1115–1129, August 2007.

[163] Ross D. Shachter. Bayes-Ball: The Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams). In *Conference on Uncertainty in Artificial Intelligence (UAI)*, 1998.

[164] Alexander Shekhovtsov, Ivan Kovtun, and Vaclav Hlavac. Efficient MRF deformation model for non-rigid image matching. *Computer Vision and Image Understanding (CVIU)*, 112(1):91–99, October 2008.

[165] Jamie Shotton, John Winn, Carsten Rother, and Antonio 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):2–23, 2009.

[166] Leonid Sigal and Michael J. Black. Measure Locally, Reason Globally: Occlusion-sensitive Articulated Pose Estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[167] Leonid Sigal, Michael Isard, Benjamin H. Sigelman, and Michael J. Black. Attractive People: Assembling Loose-Limbed Models using Non-parametric Belief Propagation. In *Advances in Neural Information Processing Systems (NIPS)*, 2003.

[168] Dheeraj Singaraju, Leo Grady, and Rene Vidal. P-brush: Continuous valued MRFs with normed pairwise distributions for image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2009.

[169] David Sontag and Tommi Jaakkola. New outer bounds on the marginal polytope. In *Advances in Neural Information Processing Systems (NIPS)*, 2007.

[170] Thad Starner, Joshua Weaver, and Alex Pentland. A wearable computer based american sign language recognizer. In *Assistive Technology and Artificial Intelligence: Applications in Robotics, User Interfaces and Natural Language Processing*, pages 84–96. Springer-Verlag, 1998.

[171] Petter Strandmark and Fredrik Kahl. Parallel and distributed graph cuts by dual decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[172] Erik B. Sudderth, Alexander T. Ihler, Michael Isard, William T. Freeman, and Alan S. Willsky. Nonparametric belief propagation. *Communications of the ACM*, 53(10):95–103, 2010.

[173] Erik B. Sudderth, Michael I. Mandel, William T. Freeman, and Alan S. Willsky. Distributed occlusion reasoning for tracking with nonparametric belief propagation. In *Advances in Neural Information Processing Systems (NIPS)*, 2004.

[174] Erik B. Sudderth, Michael I. Mandel, William T. Freeman, and Alan S. Willsky. Visual Hand Tracking Using Nonparametric Belief Propagation. *IEEE CVPR Workshop on Generative Model Based Vision*, 2004.

[175] Jian Sun, Nan-ning Zheng, and Heung-yeung Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 25(7):787–800, July 2003.

[176] Charles Sutton and Andrew McCallum. An Introduction to Conditional Random Fields. *Foundations and Trends in Machine Learning (To appear)*, 2011.

[177] Richard Szeliski. *Computer vision: algorithms and applications*. Springer-Verlag New York Inc., 2010.

[178] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten 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):1068–1080, June 2008.

[179] Marshall F. Tappen and William T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. *IEEE International Conference on Computer Vision (ICCV)*, 2003.

[180] Daniel Tarlow, Inmar E. Givoni, and Richard S. Emel. HOP-MAP: Efficient Message Passing with High Order Potentials. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010.

[181] Demetri Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 8(4):413–424, 1986.

[182] Demetri Terzopoulos and Richard Szeliski. Tracking with Kalman snakes. In *Active vision*, pages 3–20. MIT Press, 1993.

[183] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. Winston Washington, DC:, 1977.

[184] Lorenzo Torresani, Vladimir Kolmogorov, and Carsten Rother. Feature Correspondence via Graph Matching: Models and Global Optimization. In *European Conference on Computer Vision (ECCV)*, 2008.

[185] Florence Tupin, Henri Maitre, Jean-Francois Mangin, Jean-Marie Nicolas, and Eugene Pechersky. Detection of linear features in SAR images: application to road network extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 36(2):434–453, March 1998.

[186] Vijay V. Vazirani. *Approximation Algorithms*. Springer, 2001.

[187] Olga Veksler. Star Shape Prior for Graph-Cut Image Segmentation. In *European Conference on Computer Vision (ECCV)*, 2008.

[188] Olga Veksler, Yuri Boykov, and Paria Mehrani. Superpixels and supervoxels in an energy optimization framework. In *European Conference on Computer Vision (ECCV)*, 2010.

[189] Sara Vicente, Vladimir Kolmogorov, and Carsten Rother. Graph cut based image segmentation with connectivity priors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[190] Sara Vicente, Vladimir Kolmogorov, and Carsten Rother. Joint optimization of segmentation and appearance models. In *IEEE International Conference on Computer Vision (ICCV)*, 2009.

[191] George Vogiatzis, Carlos Hernández Esteban, Philip H. S. Torr, and Roberto Cipolla. Multiview stereo via volumetric Graph-Cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(12):2241–2246, December 2007.

[192] Martin J. Wainwright, Tommi Jaakkola, and Alan Willsky. Tree consistency and bounds on the performance of the max-product algorithm and its generalizations. *Statistics and Computing*, 14(2):143–166, April 2004.

[193] Martin J. Wainwright, Tommi S. Jaakkola, and Alan S. Willsky. MAP estimation via agreement on trees: Message-passing and linear programming. *IEEE Transactions on Information Theory*, 51(11):3697–3717, November 2005.

[194] Martin J. Wainwright and Michael I. Jordan. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2007.

[195] Chaohui Wang, Martin de La Gorce, and Nikos Paragios. Segmentation, ordering and multi-object tracking using graphical models. In *IEEE International Conference on Computer Vision (ICCV)*, 2009.

[196] Chaohui Wang, Olivier Teboul, Fabrice Michel, Salma Essafi, and Nikos Paragios. 3D knowledge-based segmentation using pose-invariant higher-order graphs. In *International Conference, Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2010.

[197] Chaohui Wang, Yun Zeng, Loic Simon, Ioannis Kakadiaris, Dimitris Samaras, and Nikos Paragios. Viewpoint invariant 3d landmark model inference from monocular 2d images using higher-order priors. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[198] Yair Weiss and William T. Freeman. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *IEEE Transactions on Information Theory*, 47(2):736–744, 2001.

[199] Tomás Werner. A linear programming approach to max-sum problem: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(7):1165–1179, July 2007.

[200] Tomás Werner. High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (MAP-MRF). *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[201] Tomás 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):1474–1488, August 2010.

[202] Oliver J. Woodford, Philip H. S. Torr, Ian D. Reid, and Andrew W. Fitzgibbon. Global Stereo Reconstruction under Second-Order Smoothness Priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 31(12):2115–2128, 2009.

[203] Wei Wu, Michael J. Black, Yun Gao, Elie Bienenstock, M. Serruya, A. Shaikhouni, and John P. Donoghue. Neural Decoding of Cursor Motion using a Kalman Filter. In *Advances in Neural Information Processing Systems (NIPS)*, 2002.

[204] Bo Xiang, Chaohui Wang, Jean-Francois Deux, Alain Rahmouni, and Nikos Paragios. Tagged cardiac mr image segmentation using boundary & regional-support and graph-based deformable priors. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2011.

[205] Chen Yanover, Talya Meltzer, and Yair Weiss. Linear Programming Relaxations and Belief Propagation-An Empirical Study. *The Journal of Machine Learning Research*, 7:1887–1907, 2006.

[206] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Understanding Belief Propagation and its Generalizations. In *Exploring artificial intelligence in the new millennium*, pages 239–269. Morgan Kaufmann, 2003.

[207] Yun Zeng, Chaohui Wang, Yang Wang, Xianfeng Gu, Dimitris Samaras, and Nikos Paragios. Dense non-rigid surface registration using high-order graph matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[208] Yun Zeng, Chaohui Wang, Yang Wang, Xianfeng Gu, Dimitris Samaras, and Nikos Paragios. A generic local deformation model for shape registration. Technical report, INRIA, RR-7676, July 2011.

[209] Yun Zeng, Chaohui Wang, Yang Wang, Xianfeng Gu, Dimitris Samaras, and Nikos Paragios. Intrinsic dense 3d surface tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[210] Yongmian Zhang and Qiang Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(5):699–714, May 2005.

[211] Yuhang Zhang, Richard Hartley, John Mashford, and Stewart Burn. Superpixels via Pseudo-Boolean Optimization. In *ICCV*, 2011.

[212] Song Chun Zhu and Alan 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):884–900, 1996.

Contents

1 Introduction **3**

2 Preliminaries: Graphical Models **4**
2.1 Notations . . . 4
2.2 Bayesian Networks (Directed Graphical Models) . . . 5
2.3 Markov Random Fields (Undirected Graphical Models) . . . 6
2.4 Factor Graphs . . . 8

3 MRF-based Vision Models **9**
3.1 Pairwise MRF Models . . . 9
3.1.1 Grid-like Models . . . 9
3.1.2 Part-based Models . . . 11
3.2 Higher-order MRF Models . . . 12
3.3 Conditional Random Fields . . . 13

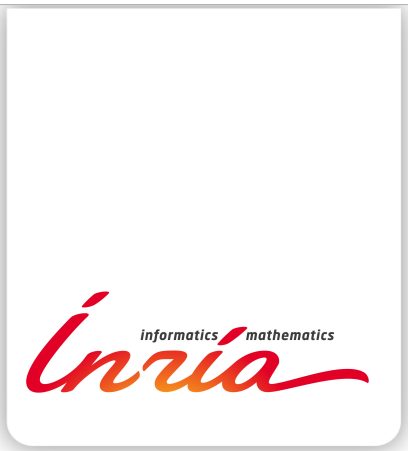
4 MAP Inference Methods for Discrete MRFs **14**
4.1 Graph Cuts and Extensions . . . 15
4.2 Belief Propagation Algorithms . . . 17
4.2.1 Belief Propagation in Tree . . . 18
4.2.2 Loopy Belief Propagation . . . 18
4.2.3 Junction Tree Algorithm . . . 19
4.3 Dual Methods . . . 19
4.3.1 Tree-reweighted Message Passing . . . 20
4.3.2 Dual Decomposition . . . 21
4.4 Inference in Higher-order MRFs . . . 22
4.4.1 Order Reduction and Graph Cuts . . . 22
4.4.2 Belief-propagation-based Methods . . . 23
4.4.3 Dual Methods . . . 23

5 Conclusion **24**

A Submodularity of MRFs **24**

B Min-sum Belief Propagation in Factor Tree **25**

C Junction Tree Algorithm **26**



RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

Parc Orsay Université
4 rue Jacques Monod
91893 Orsay Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399