Nonlinearly Constrained MRFs: Exploring the Intrinsic Dimensions of Higher-Order Cliques

Yun Zeng
Harvard University
yzeng@fas.harvard.edu

Chaohui Wang, Stefano Soatto
University of California, Los Angeles
{ch.wang, soatto}@cs.ucla.edu

Shing-Tung Yau
Harvard University
yau@math.harvard.edu

Abstract

This paper introduces an efficient approach to integrating non-local statistics into the higher-order Markov Random Fields (MRFs) framework. Motivated by the observation that many non-local statistics (e.g., shape priors, color distributions) can usually be represented by a small number of parameters, we reformulate the higher-order MRF model by introducing additional latent variables to represent the intrinsic dimensions of the higher-order cliques. The resulting new model, called NC-MRF, not only provides the flexibility in representing the configurations of higher-order cliques, but also automatically decomposes the energy function into less coupled terms, allowing us to design an efficient algorithmic framework for maximum a posteriori (MAP) inference. Based on this novel modeling/inference framework, we achieve state-of-the-art solutions to the challenging problems of class-specific image segmentation and template-based 3D facial expression tracking, which demonstrate the potential of our approach.

1. Introduction

Higher-order Markov Random Fields (MRFs) [21] have been shown to be a very powerful tool in solving many computer vision problems, largely due to their ability of representing non-local interactions/statistics [17, 26], or modeling geometric constraints that are invariant under various transformation groups [22, 23, 29]. Nevertheless, it remains challenging to even approximate the maximum a posteriori (MAP) configurations on graphs with large clique size, since most MAP inference algorithms scale exponentially with the size of the maximal clique in the graph [21, 25] (in the general case, even encoding the higher-order potentials would require exponentially large space with respect to the clique size). To overcome such a limitation, previous methods usually exploit tractable structures in higher-order cliques by representing the higher-order potentials us-
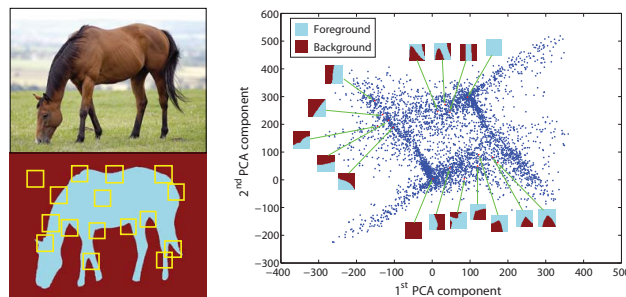
(a) An image with its seg. (b) PCA on vectorized F/B patches
Figure 1. Low dimensionality of higher order cliques.

ing simpler functions [6, 7, 9, 16, 18]. In this paper, we take a further step along this direction by exploring the intrinsic dimensions of higher-order cliques, leading to a new MRF-based modeling/inference framework which broadens the applicability of higher-order MRFs.

Our approach is largely motivated by the challenge of introducing non-local statistics into the higher-order MRF framework. On the one hand, due to the difficulty in MAP inference, most existing MRF-based models only consider low-order (*e.g.*, pairwise) potentials, resulting in biases away from their true distributions. On the other hand, many existing machine learning techniques are capable of representing the non-local statistics/priors using a small number of parameters. For instance, component analysis based approaches have been successfully applied in representing shape priors [4, 24] or image statistics [17]. An example is shown in Fig. 1. If we look at all the foreground/background patterns of 35×35 patches (Fig. 1(a)), one should be convinced that the plausible patch configurations constitute only a small fraction of the total $2^{35 \times 35}$ possibilities. For instance, principal component analysis (PCA) shows that such plausible configurations exhibits a low intrinsic dimension, as shown in Fig. 1(b). Hence the practical applicability of MRFs would be largely extended if we can design efficient MAP-MRF inference algorithms by leveraging the fact that the configurations of a higher-order clique can be represented in a low-dimensional space.

To this purpose, the first contribution of this paper is to propose *a formulation of higher-order MRFs that represents the higher-order potentials efficiently*, called *Nonlinearly Constrained MRF* (NC-MRF). By assuming each higher-order clique be represented using its local coordinates, we reformulate the MRF model by introducing additional latent variables to represent the intrinsic dimensions of the higher-order cliques. Compared to the traditional higher-order MRF formulation, NC-MRF provides a compact formalism that allows efficiently exploring the regularity and/or sparsity of the possible configurations of higher-order cliques. Moreover, it facilitates efficient inference algorithm since the formulation automatically decomposes the energy function into less coupled terms. As a consequence, it enables modeling sophisticated statistics in many vision problems.

Regarding the MAP inference, we make our second contribution by designing *an efficient algorithmic framework for the proposed NC-MRF model*. Based on a proper Lagrangian relaxation of the original problem, we achieve a new optimization framework using the primal-dual schema [19]. Similar to the dual-decomposition (DD) based inference method [1, 10], we decompose the problem into tractable slave problems which are coordinated by a master. However, rather than focusing only on the dual problem, a primal-dual pair is maintained in each iteration, allowing us to exploit the combinatorial nature of the problem. An important feature of our new framework is that it is modularized, based on the common structures of existing primal-dual methods, into components that can be customized to specific applications. Hence it provides significant flexibility in controlling the convergence rate of the algorithm, as well as the quality of the solution.

As a result, in our third contribution, we demonstrate *the potential of our modeling/inference framework in two challenging applications: class-specific segmentation and template-based* 3D *facial expression tracking*.

– In the class-specific segmentation problem, based on the fact that the foreground/background patterns of each non-local patch only span a subset of all the possible configurations, higher-order cliques are used to represent such constrained patterns learnt by PCA. By combining both the low-level color cues and the higher-order shape constraints in our NC-MRF framework, a robust solution is obtained. Our method is validated on the popular Weizmann horses data set [2].

– In the even more challenging template-based 3D facial expression tracking problem, existing approaches are either based on low-level cues with local consistency constraints ([13, 31]) or on PCA-based global shape statistics [24]. Our NC-MRF based formulation combines both low-level cues and non-local deformation constraints in the same inference framework, leading to a solution that is robust to the noisy input without losing

important details. In our experiments, we have achieved *accurate tracking results on the BU* 4D *facial expression (BU-*4*DFE) database* [27].

The rest of this paper is organized as follows: in Sec. 2 we introduce the mathematical formulation of NC-MRF and the dual of its relaxation; a new algorithmic framework for the MAP inference of NC-MRF is proposed in Sec. 3; in Sec. 4, we demonstrate two applications of our new modeling/inference framework; finally, we conclude in Sec. 5.

2. Nonlinearly constrained MRF model

Let us define a hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V}$ the vertex set and $\mathcal{E} \subset 2^{\mathcal{V}}$ the hyperedge set. Also a random variable $x_v \in \mathcal{L} \triangleq \{1, \ldots, L\}$ ($L \in \mathbb{N}^+$) is associated to each $v \in \mathcal{V}$. The MAP-MRF problem with both unary potentials ($\theta_v : \mathcal{L} \mapsto \mathbb{R}, v \in \mathcal{V}$) and higher-order potentials ($\theta_e : \mathcal{L}^{|e|} \mapsto \mathbb{R}, e \in \mathcal{E}$) can be defined as follows

$$\arg \min_{\mathbf{x} \in \mathcal{L}^{|\mathcal{V}|}} \{E(\mathbf{x}) = \sum_{v \in \mathcal{V}} \theta_v(x_v) + \sum_{e \in \mathcal{E}} \theta_e(\mathbf{x}_e)\}. \quad (1)$$

Here $\mathbf{x}_e$ represents the set of variables included in clique e.

In this paper, we represent each higher-order configuration $\mathbf{x}_e$ ($e \in \mathcal{E}$) via a mapping

$$\chi^e : \mathbf{u}_e \in \mathbb{R}^{l_e} \mapsto \mathbf{x}_e, \quad (2)$$

where $\mathbf{u}_e$ denotes the local coordinates and $l_e \in \mathbb{N}^+$. If we let $\psi_e(\mathbf{u}_e) \triangleq \theta_e(\chi^e(\mathbf{u}_e)), \forall e \in \mathcal{E}$, then problem (1) can be reformulated into the following *Nonlinearly Constrained MRF* (NC-MRF) optimization problem

$$\arg \min_{\mathbf{x}, \mathbf{u}} \quad E(\mathbf{x}, \mathbf{u}) = \sum_{v \in \mathcal{V}} \theta_v(x_v) + \sum_{e \in \mathcal{E}} \psi_e(\mathbf{u}_e) \quad (3)$$

$$\text{s.t.} \quad \chi^e_{e_i}(\mathbf{u_e}) = x_{e_i}, \qquad \forall e_i \in e \text{ and } e \in \mathcal{E},$$

where $\mathbf{x} \in \mathcal{L}^{|\mathcal{V}|}$, $\mathbf{u} \triangleq (\mathbf{u}_e)_{e \in \mathcal{E}}$, and $\chi^e_{e_i}(\cdot)$ denotes the ith component of the mapping $\chi^e(\cdot)$ (Eq. 2). Note that Eq. 3 includes the original MAP-MRF problem (1) as a special case[1]. Besides, it provides the flexibility of controlling the complexities of the higher-order potentials (*e.g.*, when $l_e \ll |e|$, the complexity of problem (3) is significantly reduced).

The optimization problem in Eq. 3 belongs to the class of Mixed Integer Nonlinear Programming (MINLP) [15]. A standard approach to solving MINLP is very similar to that of the dual-decomposition method [1], *i.e.*, *it first decomposes the original problem into a sum of sub-problems, and then solves the Lagrangian relaxation on its primal and/or dual domain*. Since problem (3) is already decomposed into terms that are coupled only by simple equality constraints, here we follow the same dual-decomposition based method in searching its MAP solutions. To this end, we first define a proper continuous relaxation of problem (3) in Sec. 2.1; then we discuss its dual problem in Sec. 2.2.

---

[1]The equivalence can be established if we let $l_e = |e|, \forall e \in \mathcal{E}$ and define: (a) $\chi^e_{e_i}(\mathbf{u}_e) = k$ if $u_{e_i} \in [k, k+1)$ and $k \in \{2, \ldots, L-1\}$; (b) $\chi^e_{e_i}(\mathbf{u}_e) = 1$ if $u_{e_i} < 2$; (c) $\chi^e_{e_i}(\mathbf{u}_e) = L$ if $u_{e_i} \geq L-1$.

2.1. A continuous relaxation of NC-MRF

Although Eq. 3 is already decomposed into a separable form, its terms $\theta_v(\cdot)$ ($v \in \mathcal{V}$) and $\psi_e(\cdot)$ ($e \in \mathcal{E}$) are not necessarily convex, leading to potentially large duality gap in its continuously relaxed problem. Following the same technique used in the LP-relaxation for MRFs [20], let us first define indicator variables $\tau_{v:j} \triangleq [x_v = j]$ [2] for each vertex $v \in \mathcal{V}$ and label $j \in \mathcal{L}$. Similarly, for the higher-order cliques we can define $\chi^e_{e_i:j}(\mathbf{u}_e) \triangleq [\chi^e_{e_i}(\mathbf{u}_e) = j]$ for each $e_i \in e$ and $j \in \mathcal{L}$. It follows that the NC-MRF optimization problem (Eq. 3) can be reformulated as

$$\arg\min_{\tau, \mathbf{u}} \quad \overline{E}(\tau, \mathbf{u}) = \sum_{v \in \mathcal{V}} \sum_{j \in \mathcal{L}} \theta_{v:j} \tau_{v:j} + \sum_{e \in \mathcal{E}} \psi_e(\mathbf{u}_e) \quad (4)$$

$$\text{s.t.} \quad \chi^e_{e_i:j}(\mathbf{u}_e) = \tau_{e_i:j}, \forall e \in \mathcal{E}, e_i \in e, j \in \mathcal{L}$$

$$\sum_{j \in \mathcal{L}} \tau_{v:j} = 1, \forall v \in \mathcal{V}$$

$$\tau_{v:j} \in \{0,1\}, \forall v \in \mathcal{V}, j \in \mathcal{L}$$

where $\tau \triangleq (\tau_{v:j})_{v \in \mathcal{V}, j \in \mathcal{L}}$, $\theta_{v:j} \triangleq \theta_v(j)$ ($\forall v \in \mathcal{V}, j \in \mathcal{L}$). A continuous relaxation of the above MINLP can be obtained by letting $\tau_{v:j} \geq 0$ ($\forall v \in \mathcal{V}, j \in \mathcal{L}$). Note that in contrast to the LP-based relaxation of the MAP-MRF problem [9, 10, 20], the functions $\psi_e(\mathbf{u}_e)$ ($e \in \mathcal{E}$) are not necessarily linear or convex. Hence it avoids introducing a very large amount of auxiliary indicator variables for all the possible configurations in the higher-order cliques. In the next section, we show that this relaxation scheme provides high flexibility in designing efficient MAP algorithms.

2.2. Dual problem and optimality conditions

Given the primal problem defined by Eq. 4, now we look at its dual problem based on Lagrangian relaxation [1]. By introducing Lagrange multipliers $\mu^e_{e_i:j}$ for each constraint $\chi^e_{e_i:j}(\mathbf{u}_e) = \tau_{e_i:j}$ ($e \in \mathcal{E}, e_i \in e, j \in \mathcal{L}$) and h_v for each constraint $\sum_{j \in \mathcal{L}} \tau_{v:j} = 1$ ($v \in \mathcal{V}$) in Eq. 4, its Lagrangian function takes the form

$$L(\tau, \mathbf{u}; \mu, \mathbf{h}) = \sum_{v \in \mathcal{V}} \sum_{j \in \mathcal{L}} \theta_{v:j} \tau_{v:j} + \sum_{e \in \mathcal{E}} \psi_e(\mathbf{u}_e) + \quad (5)$$

$$\sum_{e \in \mathcal{E}} \sum_{e_i \in e} \sum_{j \in \mathcal{L}} \mu^e_{e_i:j} (\chi^e_{e_i:j}(\mathbf{u_e}) - \tau_{e_i:j}) + \sum_{v \in \mathcal{V}} h_v (1 - \sum_{j \in \mathcal{L}} \tau_{v:j})$$

$$= \sum_{v \in \mathcal{V}} \sum_{j \in \mathcal{L}} (\theta_{v:j} - h_v - \sum_{\{e|v \in e\}} \mu^e_{v:j}) \tau_{v:j} + \quad (6)$$

$$\sum_{e \in \mathcal{E}} (\psi_e(\mathbf{u}_e) + \sum_{e_i \in e} \sum_{j \in \mathcal{L}} \mu^e_{e_i:j} \chi^e_{e_i:j}(\mathbf{u}_e)) + \sum_{v \in \mathcal{V}} h_v,$$

where $\mu \triangleq (\mu^e_{e_i:j})_{e \in \mathcal{E}, e_i \in e, j \in \mathcal{L}}$ and $\mathbf{h} \triangleq (h_v)_{v \in \mathcal{V}}$ are the Lagrange multipliers. Then the dual function for problem 4 is defined as follows

$$\text{Dual}(\mu, \mathbf{h}) \triangleq \inf_{\tau \geq 0, \mathbf{u}} L(\tau, \mathbf{u}; \mu, \mathbf{h}) \quad (7)$$

and its dual problem becomes

$$\sup_{\mu, \mathbf{h} \in \mathcal{D}} \text{Dual}(\mu, \mathbf{h}), \quad (8)$$

---

[2] $[p] = 1$ if p is true and $[p] = 0$ otherwise.

where $\mathcal{D}$ is the domain of $\text{Dual}(\cdot)$ such that $\text{Dual}(\mu, \mathbf{h}) \in (-\infty, +\infty)$. It can be proven that the following properties hold for the dual function $\text{Dual}(\cdot)$ (Eq. 7).

Proposition 1 (*Properties of the dual function*)

1. (*Weak duality*) $OPT(8) \leq OPT(4)$, where $OPT(P)$ denotes the optimal value of the problem P.

2. *The domain $\mathcal{D}$ of the dual function $\text{Dual}(\cdot)$ is convex and $\text{Dual}(\cdot)$ is concave over $\mathcal{D}$.*

3. *For the value of the dual problem to be finite, we have (otherwise by letting $\tau_{v:j} \to +\infty$, $\text{Dual}(\mu, \mathbf{h}) \to -\infty$)*

$$\theta_{v:j} - h_v - \sum_{\{e|v \in e\}} \mu^e_{v:j} \geq 0, \qquad \forall v \in \mathcal{V}, j \in \mathcal{L},$$

namely, $h_v \leq \min_{j \in \mathcal{L}} (\theta_{v:j} - \sum_{\{e|v \in e\}} \mu^e_{e_i:j})$.

4. *Following Prop. 1(3), by defining $\overline{\mathbf{h}}(\mu) = (\overline{h}_v(\mu))_{v \in \mathcal{V}}$ where $\overline{h}_v(\mu) = \min_{j \in \mathcal{L}} (\theta_{v:j} - \sum_{\{e|v \in e\}} \mu^e_{v:j})$, we have*

$$\sup_{\mu, \mathbf{h}} \text{Dual}(\mu, \mathbf{h}) = \sup_{\mu} \text{Dual}(\mu, \overline{\mathbf{h}}(\mu)), \forall (\mu, \mathbf{h}) \in \mathcal{D}$$

Hence the variables $\mathbf{h}$ can be eliminated from Eq. 8.

5. (*Subgradient*) *For any $(\lambda, \overline{h}(\lambda)) \in \mathcal{D}$, if $(\hat{\tau}, \hat{\mathbf{u}}) \in \arg\min_{\tau, \mathbf{u}} L(\tau, \mathbf{u}; \lambda, \overline{h}(\lambda))$ and $\sum_{j \in \mathcal{L}} \hat{\tau}_{v:j} = 1, \forall v \in \mathcal{V}$, the vector $g(\hat{\tau}, \hat{\mathbf{u}}) \triangleq (\chi^e_{e_i:j}(\hat{\mathbf{u}}) - \hat{\tau}_{e_i:j})_{e \in \mathcal{E}, e_i \in e, j \in \mathcal{L}}$ is a supergradient of $\text{Dual}(\cdot)$ at λ, i.e., $\forall (\mu, \overline{h}(\mu)) \in \mathcal{D}$*

$$\text{Dual}(\mu, \overline{\mathbf{h}}(\mu)) \leq \text{Dual}(\lambda, \overline{\mathbf{h}}(\lambda)) + g(\hat{\tau}, \hat{\mathbf{u}})^T (\mu - \lambda)$$

Furthermore, following the same argument in Prop. 1(3), it is not difficult to see that the following must hold true for each $v \in \mathcal{V}, j \in \mathcal{L}$.

Lemma 1 (*Complementary slackness conditions*) *For an optimal primal-dual solution $(\tau, \mathbf{u}; \mu, \mathbf{h})$, we have*

$$(a.1) \quad \tau_{v:j} > 0 \Longrightarrow h_v = \theta_{v:j} - \sum_{\{e|v \in e\}} \mu^e_{v:j}$$

$$(a.2) \quad h_v < \theta_{v:j} - \sum_{\{e|v \in e\}} \mu^e_{v:j} \Longrightarrow \tau_{v:j} = 0.$$

Finally, regarding the sufficient conditions when the problem of Eq. 4 reaches its global optimality, we have the following lemma.

Lemma 2 (*Strong agreement conditions*) *For an optimal primal-dual solution $(\tau, \mathbf{u}; \mu, \mathbf{h})$, if the following holds*

$$(b.1) \quad \mathbf{u}_e \in \arg\min_{\mathbf{u}_e} \{\psi_e(\mathbf{u}_e) + \sum_{e_i \in e} \sum_{j \in \mathcal{L}} \mu_{e_i:j} \chi^e_{e_i:j}(\mathbf{u}_e)\}$$

$$(b.2) \quad \chi^e_{v:j}(\mathbf{u}_e) = [\tau_{v:j} > 0] \Longrightarrow$$

$$h_v = \theta_{v:j} - \sum_{\{e|v \in e\}} \mu^e_{v:j}, \forall e \in \mathcal{E}, e_i \in e, j \in \mathcal{L},$$

then the optimal integer solution Eq. 3 is achieved by setting $x_v = \{j | \tau_{v:j} > 0\}, \forall v \in \mathcal{V}$. The uniqueness of x_v is guaranteed by the definition of χ^e (Eq. 2).

The correctness of Lemma 2 can be proven by checking the duality gap becomes zeros if the strong agreement conditions are met. Based on the above analysis of the optimality conditions, now we are ready for designing efficient optimization algorithms.

3. MAP inference via primal-dual schema

Towards efficient MAP inference for the NC-MRF model, we propose an algorithmic framework based on *primal-dual schema*, referring to the class of approximation algorithms [19] that iteratively solve the problem by maintaining both a primal integral solution and a feasible dual solution in each iteration. As noted in [19], the advantage of primal-dual schema is that they allow us to explore the combinatorial structures of the problems and thus often achieve good runtime (*e.g.*, the FastPD algorithm [11]).

The basic ideas of our design are as follows:

1. In each iteration, we maintain both an integer solution $(x_v)_{v \in \mathcal{V}}$ ($x_v \in \mathcal{L}$), a continuous primal solution τ and a dual solution μ. Given a dual solution μ, the primal solution τ is chosen such that both Lemma 1 and the condition $\sum_{j \in \mathcal{L}} \tau_{v;j} = 1$ are met, which restricts us to assign nonzero values only to $\tau_{v;j}$'s with label j from

$$\text{Sol}_v(\mu) \triangleq \{j | \overline{h}_v(\mu) = \theta_{v;j} - \sum_{e|v \in e} \mu_{v;j}^e, j \in \mathcal{L}\}. \quad (9)$$

   $\text{Sol}_v(\mu)$ is guaranteed to be nonempty if we consider the dual function $\text{Dual}(\mu, \overline{\mathbf{h}}(\mu))$ defined in Prop. 1 (4). Intuitively, $\tau_{v;j}$ denotes the probability for label j to be the primal solution of the integer programming problem of Eq. 4. Accordingly, each x_v is also chosen from those labels with non-zero probability, *i.e.*, $\text{Sol}_v(\mu)$.

2. Following 1, the first term in Eq. 6 disappears. The dual problem then becomes the following

$$\max_{\mu} \{\text{Dual}^I(\mu) = \sum_{e \in \mathcal{E}} s_e(\mu_e) + \sum_{v \in \mathcal{V}} \overline{h}_v(\mu)\}. \quad (10)$$

   Here each subproblem $s_e(\cdot)$ is defined by

$$s_e(\mu_e) = \min_{\mathbf{u}_e} \{\psi_e(\mathbf{u}_e) + \sum_{e_i \in e} \sum_{j \in \mathcal{L}} \mu_{e_i:j} \chi_{e_i:j}^e(\mathbf{u}_e)\} \quad (11)$$

   and $\mu_e \triangleq (\mu_{e_i:j}^e)_{e_i \in e, j \in \mathcal{L}}$. The advantage of maintaining a primal solution in each iteration, is that each subproblem can also compute its solution based on current primal solution $\mathbf{x}$, by solving a proximal problem

$$\min_{\mathbf{u}_e} \{\psi_e(\mathbf{u}_e) + \sum_{e_i \in e} \sum_{j \in \mathcal{L}} \mu_{e_i:j} \chi_{e_i:j}^e(\mathbf{u}_e) + d(\chi^e(\mathbf{u}_e), \mathbf{x})\},$$

   where $d(\cdot, \cdot)$ denotes a similarity measure between the solution proposed by the subproblem and the current integer primal solution. This scheme will encourage agreement among different subproblems [14].

3. Given the solution proposed by each subproblem, the dual variables μ are updated to optimize the dual function by the dual ascent algorithm using its subgradient.

4. After the dual ascent update, the algorithm stops if the dual objective function can no longer be improved. Otherwise, a new primal solution x_v ($v \in \mathcal{V}$) is selected from the updated set $\text{Sol}_v(\mu)$ (Eq. 9). Therefore, if Sol_v contains more than one elements, a consensus rule should be specified to decide a unique x_v.

By following the above design rules, it is guaranteed that the final solution is in the set $\text{Sol}_v(\mu), v \in \mathcal{V}$. Our new framework provides the flexibility in the design of dual ascent algorithms (Sec. 3.1) and selection of primal solutions (Sec. 3.2), which can be decided according to the need of applications. An outline of the above primal-dual (PD) strategy for the MAP inference is summarized in Alg. 1.

Algorithm 1: Outline of the PD strategy for NC-MRF

Initialization: set $t = 0$, $\mu^t = 0$,
$x_v^0 = \arg \min_j \theta_{v:j}, \forall v \in \mathcal{V}$ (break ties arbitrarily)
Repeat
    Solve each subproblem (Eq. 11) with solution $\{\chi^e(\mathbf{u}_e^t)\}$;
    Dual ascent – update μ^t (Sec. 3.1);
    Primal consensus – update $(x_v)_{v \in \mathcal{V}}$ and τ^t (Sec. 3.2);
    $t \leftarrow t + 1$;
Until no update of μ^t can improve the lower bound

3.1. Design of dual ascent algorithms

Now let us look at strategies to improve the lower bound in the dual domain. From Prop. 1(5), for each $e \in \mathcal{E}, e_i \in e, j \in \mathcal{L}$, the subgradient update is represented by

$$\mu_{e_i:j}^{e,t+1} = \mu_{e_i:j}^{e,t} + \alpha^t (\chi_{e_i:j}^e(\mathbf{u}_e^t) - \tau_{e_i:j}^t), t = 0, 1, \ldots \quad (12)$$

where $\alpha^t > 0$ denotes the step size at iteration t. Intuitively, this update rule simply states that if the current primal solution ($\tau_{e_i:j}^t$) is different from the solution ($\chi_{e_i:j}^e(\mathbf{u}_e^t)$) proposed by the subproblem $s_e(\cdot)$ on label j, the value of $\mu_{e_i:j}^e$ should be increased ($\chi_{e_i:j}^e(\mathbf{u}_e^t) > 0$ and $\tau_{e_i:j}^t = 0$) or decreased ($\chi_{e_i:j}^e(\mathbf{u}_e^t) = 0$ and $\tau_{e_i:j}^t > 0$)[3].

However, as noted in [5], one issue with directly optimizing the dual problem of Eq. 7 is that the convergence can be very slow when the number of subproblems is large. Intuitively, the more subproblems we have, the more difficult to reach an agreement. A standard approach to get around this is to define a surrogate/proximal function that relaxes the original objective function [5, 11, 14]. In fact, a very straightforward surrogate function can be simply defined by

$$\text{Dual}^{II}(\mu) = \sum_{v \in \mathcal{V}} \overline{h}_v(\mu). \quad (13)$$

---

[3]Note that this dual ascent strategy underlies many dual optimization based MRF inference algorithms [8, 10, 11, 25]. When the subproblems $s_e(\cdot)$ are formulated into LP, the dual objective of Eq. 10 can be improved using algorithms like dual averaging [8, 10] or max-sum diffusion [25].

It is a good approximation to $\text{Dual}^I(\cdot)$ in Eq. 10 when the dual function $s_e(\cdot)$ is relatively flat (*i.e.*, given μ_e, it is always possible to find a primal solution $\mathbf{u}_e$ that achieves the same minimum) – this is the case when the higher-order potentials $\psi_e(\mathbf{u}_e)$ represents only configurational constraints.

In order to optimize the dual objective of Eq. 13, we only have to solve the following problem

$$\arg\max_{\alpha^t \geq 0} \text{Dual}^{II}(\mu^{t+1}(\alpha^t)), \qquad (14)$$

where $\mu^{t+1}(\alpha^t)$ denotes the subgradient update. To facilitate fast convergence, we consider the subgradient update defined by $\mu_{e_i:j}^{e,t+1} = \mu_{e_i:j}^{e,t} + \alpha^t(\chi_{e_i:j}^e(\mathbf{u}_e^t) - \tau_{e_i:j}^t) * [\chi_{e_i:j}^e(\mathbf{u}_e^t) * \tau_{e_i:j}^t = 0]$, which guarantees that the objective function Dual^{II} will first monotonically increase then decrease as α^t increases from zero [28]. It can be shown that the problem of Eq. 14 can be solved by an $O(|\mathcal{V}||\mathcal{E}|\log(|\mathcal{V}||\mathcal{E}|))$ algorithm, and converges in $O(\frac{1}{\epsilon})$ iterations for an ϵ-accurate solution [28]. Therefore, the dual optimization algorithm only scales in the order of $n\log n$, where n is the number of cliques in the graph.

3.2. Design of primal consensus rules

Let us recall the process of the primal-dual update in the tth iteration of Alg. 1: (1) for each node $v \in \mathcal{V}$, after each subproblem is solved based on current primal-dual solution, a new solution is proposed by each clique e with $v \in e$; (2) if there are proposed solutions not in the current set Sol_v (Eq. 9), then the dual optimization algorithm updates μ based on the newly proposed solutions (Sec. 3.1); (3) accordingly, the set $\text{Sol}_v(\mu^t)$ is updated as well, with some labels included and some excluded. Eventually, the dual objective function will no longer be improved and a unique solution is determined from $\text{Sol}_v(\mu^t)$. If there are cliques that disagree with the final solution, their proposed solutions will simply be disregarded. Therefore, the role of the consensus rule is to guarantee that the final solution selected *reflects the agreement of the majority*. In practice, which strategy to choose is application dependent. Here we propose a general consensus rule as follows.

The idea is to assign a weight to each label, denoted by $w_{v;j}, j \in \mathcal{L}$, which is initialized to 1. The weights are updated in each iteration as follows: if the clique e proposes a solution $j \in \mathcal{L}$ for node v, the weight can be updated by $w_{v;j} \leftarrow w_{v;j} * (1 + k_e^v)$, where $k_e^v \geq 0$ denotes the importance of the solution proposed by clique e for node v. k_e^v can be either defined as a constant or learnt from training data. Intuitively, when there are multiple choices in the set $\text{Sol}_v(\mu)$, larger weights are given for those labels with larger number of cliques that agree on them. Hence a primal integral solution can be chosen from the set $\text{Sol}_v(\mu)$ with largest weight (tie is broken arbitrary). Given the weighting system defined above, we can simply determine the primal solution for τ as

$$\tau_{v;j} = \begin{cases} w_{v;j}/\sum_{j \in \text{Sol}_v} w_{v;j} & \text{if } j \in \text{Sol}_v, \\ 0 & \text{otherwise} \end{cases},$$

which is used for the subsequent dual optimization in the next iteration.

At this point, we have presented a complete algorithmic framework for solving the NC-MRF inference problem of Eq. 4 based on the common structure of exiting dual-optimization based approaches (*e.g.*, [8, 9, 10, 25]). Since it is unlikely (unless $\mathcal{P} = \mathcal{NP}$) that any algorithm can obtain a guaranteed approximate solution to the general MRF optimization problem of Eq. 1 [3], our approach provides an alternative way to get around the boundary in computation and allows us to pay more attention to the model's efficacy, whose benefits will be demonstrated in the next section.

4. Applications

The major advantage of our new modeling/inference framework lies in its ability to encode rich non-local statistics in an efficient manner. Thus, it provides a very useful tool for developing advanced graph-based algorithms for numerous challenging problems in computer vision. To demonstrate its potential, we present two applications: class-specific image segmentation and template-based 3D facial expression tracking. In both cases, we consider the dual objective defined by Eq. 13 ($\epsilon = 10^{-3}$) and set $k_e^v = 0.05$ for the weight update described in Sec. 3.2.

4.1. Class-specific segmentation with shape priors

Given an image I, the graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is constructed by associating each pixel in I with a node $v \in \mathcal{V}$. The binary label set $\mathcal{L} \triangleq \{F = 1, B = 0\}$ for each node v consists of foreground (F) and background (B). The higher-order clique set $\mathcal{E}$ consists of all the $p \times p$ patches in the image.

An intrinsic representation of the higher-order cliques can be obtained by the PCA model learnt from all the possible F/B patterns of $p \times p$ patches (Fig. 1). Thus for the mapping function $\chi^e(\mathbf{u}_e), e \in \mathcal{E}$, we define $\mathbf{u}_e$ as the parameters for the PCA components that encode the shape priors. By choosing the first l components in the learnt PCA model, the configuration of the higher-order clique corresponding to each patch can be defined as

$$\chi^e(\mathbf{u}_e) = \Pi_{\{F,B\}^{p \times p}}(\mathbf{s}_0 + \sum_{i=1}^{l} u_i^e \mathbf{s}_i). \qquad (15)$$

Here s_i ($i = 0, 1, \ldots l$) are the learnt shape bases from PCA and $\Pi_{\{F,B\}^{p \times p}}(\cdot)$ denotes the mapping from the vectorized shape space to the binary F/B patch.

As commonly used in MRF-based segmentation approaches, the unary term is defined as follows

$$\theta_v(x_v) = \begin{cases} -\log(\Pr(I_v|\mathbf{H}^f)) & \text{if } x_v = F \\ -\log(\Pr(I_v|\mathbf{H}^b)) & \text{otherwise} \end{cases} \qquad (16)$$
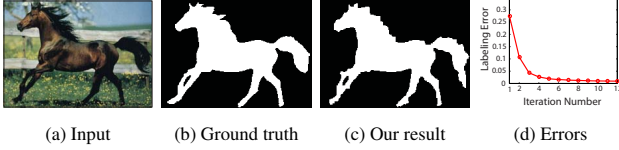
(a) Input   (b) Ground truth   (c) Our result   (d) Errors

Figure 2. An example of segmentation result, with the labeling error (defined in the text) at each iteration (d).

where $\mathbf{H}^f$ and $\mathbf{H}^b$ denote the photometric/color model for the foreground and background, respectively. We use the standard Gaussian Mixture Models (GMM) with learned parameters from training images in our experiments.

Given the unary potential of Eq. 16, we can define the potential for each higher-order clique $e \in \mathcal{E}$ as

$$\psi(\mathbf{u}_e) = \sum_{e_i \in e} \theta_{e_i}(\chi^e_{e_i}(\mathbf{u}_e)) = \sum_{e_i \in e} \sum_{j \in \mathcal{L}} \theta_{e_i;j} \chi^e_{e_i;j}(\mathbf{u}_e), \quad (17)$$

which constrains the configuration must be within the learnt prior patterns while conforming to the observation. By substituting Eq. 17 into Eq. 11, each subproblem now becomes

$$s_e(\mu_e) = \min_{\mathbf{u}_e} \sum_{e_i \in e} \sum_{j \in \mathcal{L}} (\theta_{e_i;j} + \mu_{e_i;j}) \chi^e_{e_i;j}(\mathbf{u}_e). \quad (18)$$

To obtain its solution, we first estimate its initial binary solution by ($\forall e_i \in e$)

$$\text{sol}^e(e_i) = \begin{cases} F & \text{if } \theta_{e_i;F} + \mu_{e_i;F} \leq \theta_{e_i;B} + \mu_{e_i;B} \\ B & \text{otherwise} \end{cases} \quad (19)$$

In case it is not in the allowed configuration, its projection into the allowed shape space (Eq. 15) can be easily obtained by taking the inner product $u^e_i = \langle s_i, \text{sol}^e - s_0 \rangle$.

Note that the PCA reconstruction given by $\mathbf{s}_0 + \sum_{i=1}^{l} u^e_i \mathbf{s}_i$ outputs a $p \times p$ patch with continuous values. Special care should be taken in defining the mapping $\Pi_{\{F,B\}^{p \times p}}(\cdot)$ (Eq. 15), which must guarantee the plausible F/B configuration space conforms to the shape prior. In the case of horse segmentation problem as illustrated in Fig. 1, the following topological constraints are obtained from the the learnt data: (a) each connected foreground region must be connected to one of the patch's four boundaries; (b) there is at most one hole in the foreground region; (c) each patch has at most three connected foreground regions. These conditions can be guaranteed by a binary search for the optimal threshold that separates the foreground from the background (with a complexity of only $O(p^2 \log k)$, where $k = 100$ is the discretization level in the continuous space).

Results We evaluate our method on the popular Weizmann horses data set [2] which consists of 328 images of horses with various backgrounds. The data is randomly split into $1/3$ for training and $2/3$ for testing. The patch size is chosen to be 35×35 and we use the first 300 PCA components (*i.e.*, $l = 300$ in Eq. 15). The labeling error is measured by the percentage of mislabeled pixels in the

(a) Example of results with small errors ($< 1\%$)

(b) Example of results with large errors ($> 10\%$)

Figure 3. The results on the Weizmann horses data set [2].

segmented images, as used in [12]. By simply imposing the topological constraints for each patch as discussed above, we have achieved 95.34% classification accuracy. A further improvement is made by constraining the smoothness of the F/B boundaries for each patch, leading to an 96.77% accuracy, which outperforms methods based on learning high-level structures [12]. The average running time for the test images is $1.26s$. The average number of iterations is 42. Fig. 2 shows an example of horse segmentation with the labeling error plotted for each iteration. Fig. 3 shows representative results with low and high errors.

4.2. Template-based 3D facial expression tracking

We next look at the problem of template-based 3D surface tracking. Similar to [31], a template $\mathcal{T}_0$, *i.e.*, a triangulated mesh consisting of a number of points $(v_1^0, v_2^0, \ldots, v_n^0)$, is given at the first frame of a sequence $\{\mathcal{M}^0, \ldots, \mathcal{M}^m\}$ of 3D surface data with texture information. The goal is to infer the trajectory of the template ($\mathcal{T}_t$) in each subsequent frame $t = 1, 2, \ldots, $ n. This can be represented probabilistically by

$$\arg \min_{\mathcal{T}_i, i=1,\ldots,m} -\log \Pr(\mathcal{T}_1, \ldots, \mathcal{T}_n | \mathcal{T}_0, \mathcal{M}^0, \ldots, \mathcal{M}^m). \quad (20)$$

The MRF model allows us to encode a proper probability $\Pr(\cdot)$ for Eq. 20. This said, let us define a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V}$ the node set on the template and $\mathcal{E}$ the patch set (Fig. 4 (a)). For each node $v \in \mathcal{V}$, the set of candidate 3D positions at time t is obtained by the efficient sampling method described in [31]. We assume the deformation of each each patch $e \in \mathcal{E}$ lies in a subspace which can be represented by continuous variables $\mathbf{u}_e$. Let $\chi^e(\mathbf{u}_e)$ denote the mapping from the 3D locations to the discrete labels for each of e's nodes and $\theta_v(x_v)$ denote the unary potential that measures the similarity for each correspondence based on texture/geometry information, the higher-order potential $\psi(\mathbf{u}_e)$ can then be similarly defined in the form of Eq. 17. We refer the reader to [31] for the detailed definition of the singleton potentials $\theta_v(\cdot)$. In the following, we only discuss the plausible space of the higher-order configuration $\chi^e(\mathbf{u}_e)$.
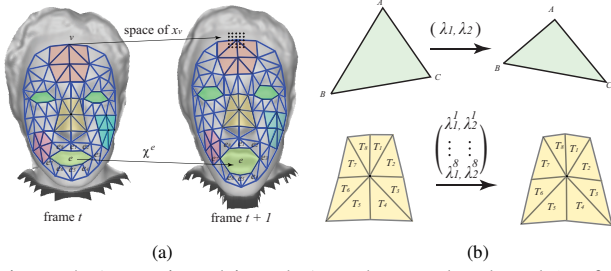
Figure 4. (Best viewed in color) In the template-based 3D face tracking problem, a template is constructed for the first frame in a 3D sequence and the locations of its nodes are inferred in the subsequent frames. Our NC-MRF formulation partitions the template into overlapping patches (colored in (a)), whose deformations are constrained by its distortions represented by CDCs [30] (b).

Representation of deformation space The challenge of deformation modeling lies in the characterization of non-rigid motion. Here, we employ the Canonical Distortion Coefficients (CDCs) proposed in [30] for representing non-rigid deformation. As shown in Fig. 4(b) top, the CDCs are defined for each triangular facet that represent the distortions along its two principle directions, denoted by $(\lambda_1, \lambda_2) \in \mathbb{R} \times \mathbb{R}$ (*e.g*, if $\lambda_1 = \lambda_2 = 1$, the deformation is rigid), which can be efficiently computed in closed form [30]. In order to apply CDCs to representing the deformation of a patch $e \subset 2^{\mathcal{V}}$, we consider a triangulation of the nodes in e, denoted by $\mathcal{F}_e \subset e \times e \times e$. Since the deformation of each triangle $f \in \mathcal{F}_e$ can be represented using CDCs, the non-rigid deformation of the whole patch can then be represented by the vector $d^e = (\lambda_1^f, \lambda_2^f)_{f \in \mathcal{F}_e}, d^e \in \mathbb{R}^{|\mathcal{F}_e| \times 2}$ (Fig. 4(b) bottom). In this way, a prior deformation space can be represented as $d(\alpha) = d_0 + \sum_{i=1}^{l} \alpha_i d_i$, and the intrinsic parameters for a patch e are simply $\mathbf{u}_e = \alpha^e$ where $\alpha^e \equiv (\alpha_i^e)_{i=1}^l$ are the coefficients for PCA. Here we select the number of components $l = 5 \sim 10$, depending on the size of the patch.

Subproblem optimization Given the above representation of deformation space, each subproblem e then becomes searching for the optimal $\mathbf{u}_e$ given current dual variables μ_e. To facilitate fast computation, similar to our approach to the image segmentation problem (Sec. 4.1), we obtain an initial solution by solving $\text{sol}^e(e_i) = \arg \min_j \{\theta_{e_i;j} + \mu_{e_i;j}\}, \forall e_i \in e$, which also gives us an initial correspondence for each node in e. A deformation vector d^e is then computed based on this initial correspondence. It follows that the PCA coefficients can be sufficiently computed by $\alpha_i^e = \langle d^e, d_i^e \rangle$, where d_i^e is the ith PCA basis. Given the desired CDCs of the patch, the optimal deformation of the patch can be solved by employing the higher-order MRF optimization proposed in [30], providing the final correspondence for each node. Note that the optimization involves significantly smaller graph size and usually it can be solved exactly due to small tree-width of the subproblem.
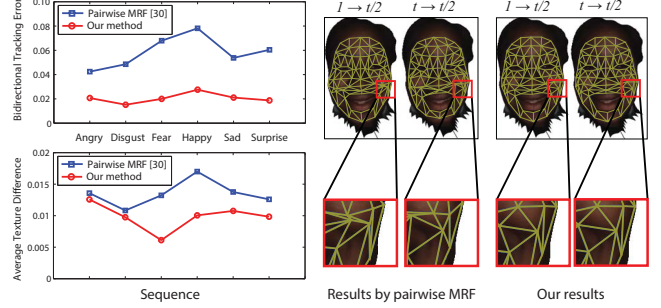
(a) Two evaluations for data F004   (b) Bidirectional results for sequence Happy

Figure 5. (Best viewed in color) A comparison between our method and the pairwise MRF approach [31] for a difficult case with unreliable texture information. (a) shows error evaluations based on Average Texture Differences (ATD) between the current frame and the first frame (down), and the Bidirectional Tracking Error (BTE) proposed in this paper (up). BTE is able to reflect the consistency in structure as illustrated in (b).

Data set and experimental setting Similar to [24], we train our model from the tracking results obtained from data with high quality geometry and texture, which are captured using the high resolution 3D scanning system described in [31]. Our algorithm is tested on the BU-4DFE database [27], consisting of 101 different persons (58 females and 43 males) each with 6 different expressions (angry, disgust, fear, happy, sad, surprise) and around 100 frames/expression. A generic template with 74 nodes (Fig. 4(a)) is first semi-automatically attached to the first frame of each sequence, and then tracked in the subsequent frames. We choose 62 overlapping patches with various sizes ($6 \sim 20$) for the higher-order deformation constraints. For each node, we sample $|\mathcal{L}| = 64$ matching candidates in each frame. Fig. 6 shows some representative tracking results [28] on the BU database.

Quantitative evaluation Quantitatively evaluating the accuracy of the tracking results is challenging since the dense correspondence ground truth are usually unavailable. We first measure the texture differences between correspondences as in [31], leading to an improvement (Fig. 5 (a)). Meanwhile, we observe that such a measurement is unable to reflect the consistency in global structure (Fig. 5 (b)). Thus, we perform an additional quantitative evaluation as follows: (1) given a sequence with t frames, we first obtain the tracking result at frame t by directly tracking it from frame 1; (2) then we track the whole sequence in two directions: forward ($1 \rightarrow t$) and backward ($t \rightarrow 1$); (3) let us denote by $p_{v;t/2}^{\rightarrow}(p_{v;t/2}^{\leftarrow}) \in \mathbb{R}^3$ the 3D position of node v at frame $t/2$ computed by direction $1 \rightarrow t$ ($t \rightarrow 1$), $\mathcal{V}$ and $\mathcal{P}$ the sets of all nodes and edges in the template, respectively, then the tracking error, called the *Bidirectional Tracking Error* (BTE), for a sequence with t frames is defined by

$$\frac{\sum_{v \in \mathcal{V}} |p_{v;t/2}^{\rightarrow} - p_{v;t/2}^{\leftarrow}|}{\min\{\sum_{(v,w) \in \mathcal{P}} |p_{v;t/2}^{\rightarrow} - p_{w;t/2}^{\rightarrow}|, \sum_{(v,w) \in \mathcal{P}} |p_{v;t/2}^{\leftarrow} - p_{w;t/2}^{\leftarrow}|\}|\}}$$

Figure 6. Facial expression tracking results on BU database [27].

Intuitively, the BTE gives us a good quantitative measure of the stability (amount of drifts) of the tracking results. For all the 101×6 sequences we tested, the average BTE is 0.0103. For comparison, we randomly pick 10 sequences from the BU database and evaluate the errors obtained by the latest pairwise MRF based method [31]. It turns out that the smallest BTE achieved by [31] is 0.0285 and our method achieves significantly smaller errors for every case, although a slight overhead is added in optimizing the higher-order cliques ($\sim$ 1s / frame for 62 cliques on CPU).

5. Conclusion

We have proposed a new representation of higher-order MRFs, namely NC-MRFs, by specifically dealing with the intrinsic dimensions of higher-order cliques. NC-MRFs provide a compact representation of higher-order MRFs by introducing additional latent variables to represent the plausible configurations of higher-order cliques, and allow flexible design of efficient inference algorithms by decomposing the energy function into less coupled terms. Notably, we have achieved state-of-the-art results for the challenging problems of class-specific image segmentation and template-based 3D facial expression tracking. The proposed modeling/inference framework provides a practical way of incorporating sophisticated higher-order constraints into graphical models, which has promising potential of solving numerous challenging problems in computer vision.

References

[1] D. P. Bertsekas. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[2] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In *ECCV*, pages 109–124, 2002.

[3] V. Chandrasekaran, N. Srebro, and P. Harsha. Complexity of inference in graphical models. In *UAI*, 2008.

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models–their training and application. *CVIU*, 61(1):38–59, 1995.

[5] V. Jojic, S. Gould, and D. Koller. Accelerated dual decomposition for MAP inference. In *ICML*, 2010.

[6] P. Kohli and M. P. Kumar. Energy minimization for linear envelope MRFs. In *CVPR*, pages 1863–1870, 2010.

[7] P. Kohli, M. P. Kumar, and P. H. Torr. P3 & beyond: Move making algorithms for solving higher order functions. *IEEE TPAMI*, 31:1645–1656, 2009.

[8] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *TPAMI*, 28(10):1568–1583, 2006.

[9] N. Komodakis and N. Paragios. Beyond pairwise energies: Efficient optimization for higher-order MRFs. In *CVPR*, 2009.

[10] N. Komodakis, N. Paragios, and G. Tziritas. MRF energy minimization and beyond via dual decomposition. *IEEE TPAMI*, 33:531–552, 2011.

[11] N. Komodakis, G. Tziritas, and N. Paragios. Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies. *CVIU*, 112(1):14–29, 2008.

[12] A. Levin and Y. Weiss. Learning to combine bottom-up and top-down segmentation. *IJCV*, 81(1):105–118, 2009.

[13] H. Li, R. W. Sumner, and M. Pauly. Global correspondence optimization for non-rigid registration of depth scans. In *SGP*, 2008.

[14] A. F. T. Martins, M. A. T. Figueiredo, P. M. Q. Aguiar, N. A. Smith, and E. P. Xing. An augmented lagrangian approach to constrained MAP inference. In *ICML*, 2011.

[15] I. Nowak. *Relaxation and Decomposition Methods for Mixed Integer Nonlinear Programming*. Birkhäuser, 2005.

[16] B. Potetz and T. S. Lee. Efficient belief propagation for higher-order cliques using linear constraint nodes. *CVIU*, 112:39–54, 2008.

[17] S. Roth and M. J. Black. Fields of experts. *IJCV*, 82(2):205–229, Apr. 2009.

[18] C. Rother, P. Kohli, W. Feng, and J. Jia. Minimizing sparse higher order energy functions of discrete variables. In *CVPR*, 2009.

[19] V. V. Vazirani. *Approximation Algorithms*. Springer-Verlag, 2001.

[20] M. Wainwright, T. Jaakkola, and A. Willsky. MAP estimation via agreement on trees: Message passing and linear-programming approaches. *IEEE Transactions on Information Theory*, 51(11):3697–3717, 2005.

[21] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in ML*, 2008.

[22] C. Wang, O. Teboul, F. Michel, S. Essafi, and N. Paragios. 3D knowledge-based segmentation using pose-invariant higher-order graphs. In *MICCAI*, 2010.

[23] C. Wang, Y. Zeng, L. Simon, I. Kakadiaris, D. Samaras, and N. Paragios. Viewpoint invariant 3D landmark model inference from monocular 2D images using higher-order priors. In *ICCV*, 2011.

[24] T. Weise, S. Bouaziz, H. Li, and M. Pauly. Realtime performance-based facial animation. *ACM Trans. Graph.*, 30(4), 2011.

[25] T. Werner. Revisiting the linear programming relaxation approach to Gibbs energy minimization and weighted constraint satisfaction. *TPAMI*, 32:1474–1488, 2010.

[26] O. J. Woodford, C. Rother, and V. Kolmogorov. A global perspective on MAP inference for low-level vision. In *ICCV*, 2009.

[27] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3D dynamic facial expression database. In *Auto. Face and Gesture Rec.*, 2008.

[28] Y. Zeng, C. Wang, S. Soatto, and S.-T. Yau. *Nonlinearly Constrained MRFs: Exploring the Intrinsic Dimensions of Higher-Order Cliques – Supplementary Material*, 2013.

[29] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios. Dense non-rigid surface registration using high-order graph matching. In *CVPR*, 2010.

[30] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios. A Generic Local Deformation Model for Shape Registration. Technical report, INRIA, 2011.

[31] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios. Intrinsic dense 3D surface tracking. In *CVPR*, 2011.