$\mathcal{P}^3$ & Beyond: Solving Energies with Higher Order Cliques

Pushmeet Kohli M. Pawan Kumar Philip H. S. Torr
Oxford Brookes University, UK
{pushmeet.kohli, pkmudigonda, philiptorr}@brookes.ac.uk
http://cms.brookes.ac.uk/research/visiongroup/

Abstract

In this paper we extend the class of energy functions for which the optimal α-expansion and $\alpha\beta$-swap moves can be computed in polynomial time. Specifically, we introduce a class of higher order clique potentials and show that the expansion and swap moves for any energy function composed of these potentials can be found by minimizing a submodular function. We also show that for a subset of these potentials, the optimal move can be found by solving an st-mincut problem. We refer to this subset as the $\mathcal{P}^n$ Potts model.

Our results enable the use of powerful move making algorithms i.e. α-expansion and $\alpha\beta$-swap for minimization of energy functions involving higher order cliques. Such functions have the capability of modelling the rich statistics of natural scenes and can be used for many applications in computer vision. We demonstrate their use on one such application i.e. the texture based video segmentation problem.

1. Introduction

In recent years discrete optimization has emerged as an important tool in solving Computer Vision problems. This has primarily been the result of the increasing use of energy minimization algorithms such as graph cuts [4, 12], tree-reweighted message passing [11, 25] and variants of belief propagation (BP) [18, 26]. These algorithms allow us to perform approximate inference (i.e. obtain the MAP estimate) on graphical models such as Markov Random Fields (MRF) and Conditional Random Fields (CRF) [15].

α-expansion and $\alpha\beta$-swap are two popular *move* making algorithms for approximate energy minimization which were proposed in [4]. They are extremely efficient and have been shown to produce good results for a number of problems [23]. These algorithms minimize an energy function by starting from an initial labelling and making a series of changes (moves) which decrease the energy iteratively. Convergence is achieved when the energy cannot be minimized further. At each step the *optimal* move (i.e. the move decreasing the energy of the labelling by the most amount) is computed in polynomial time. However, this can only be done for a certain class of energy functions.

Boykov *et al.* [4] provided a characterization of clique potentials for which the optimal moves can be computed by solving an st-mincut problem. However, their results were limited to potentials of cliques of size at most two. We call this class of energy functions $\mathcal{P}^2$. In this paper we provide the characterization of energy functions involving higher order cliques i.e. cliques of sizes 3 and beyond for which the optimal moves can be computed in polynomial time. We refer to the class of functions defined by cliques of size at most n as $\mathcal{P}^n$. It should be noted that this class is different from the class $\mathcal{F}^n$ of energy functions which involve only binary random variables [7, 12].

Higher order cliques Most energy minimization based methods for solving Computer Vision problems assume that the energy can be represented in terms of unary and pairwise clique potentials. This assumption severely restricts the representational power of these models making them unable to capture the rich statistics of natural scenes [16].

Higher order clique potentials have the capability to model complex interactions of random variables and thus could overcome this problem. Researchers have long recognized this fact and have used higher order models to improve the expressive power of MRFs and CRFs [16, 20, 21]. The initial work in this regard has been quite promising and higher order cliques have been shown to improve results. However their use has been quite limited due to the lack of efficient algorithms for minimizing the resulting energy functions.

Traditional inference algorithms such as BP are quite computationally expensive for higher order cliques. Lan *et al.* [16] recently made some progress towards solving this problem. They proposed approximation methods for BP to make efficient inference possible in higher order MRFs. However their results indicate that BP only gave comparable results to naïve gradient descent. In contrast, we provide a characterization of energy functions defined by cliques of size 3 ($\mathcal{P}^3$) or more ($\mathcal{P}^n$) which can be solved using powerful move making algorithms such as α-expansion and $\alpha\beta$-swaps. We prove that the optimal α-expansion and $\alpha\beta$-swap moves for this class of functions can be computed in polynomial time. We then introduce a new family of higher order potential functions, referred to as the $\mathcal{P}^n$ Potts model, and show that the optimal α-expansion and $\alpha\beta$-swap moves for them can be computed by solving an st-mincut problem. It should be noted that our results are a generalization of the class of energy functions specified by [4].

Outline of the Paper In section 2, we provide the notation and discuss the basic theory of energy minimization and submodular functions. Section 3 describes the α-expansion and $\alpha\beta$-swap algorithms. Further, it provides

constraints on the pairwise potentials which guarantee computation of the optimal move in polynomial time. In section 4, we generalize this class to $\mathcal{P}^n$ functions. We also show that the optimal moves for a sub-class of these functions, i.e. the $\mathcal{P}^n$ Potts model, can be computed by solving an st-mincut problem. This enables us to address the texture based segmentation problem (see section 5). We conclude by listing some Computer Vision problems where higher order clique potentials can be used.

2. Preliminaries

Consider a random field $\mathbf{X}$ defined over a lattice $\mathcal{V} = \{1, 2, \ldots, N\}$ with a neighbourhood system $\mathcal{N}$. Each random variable $X_i \in \mathbf{X}$ is associated with a lattice point $i \in \mathcal{V}$ and takes a value from the label set $\mathcal{L} = \{l_1, l_2, \ldots, l_k\}$. Given a neighborhood system $\mathcal{N}$, a clique c is specified by a set of random variables $\mathbf{X}_c$ such that $\forall i, j \in c, i \in \mathcal{N}_j$ and $j \in \mathcal{N}_i$, where $\mathcal{N}_i$ and $\mathcal{N}_j$ are the sets of all neighbours of variable X_i and X_j.

Any possible assignment of labels to the random variables will be called a *labelling* (denoted by $\mathbf{x}$). In other words, $\mathbf{x}$ takes values from the set $\mathbf{L} = \mathcal{L}^N$. The posterior distribution $\Pr(\mathbf{x}|\mathbf{D})$ over the labellings of the random field is a *Gibbs* distribution if it can be written in the form:

$$\Pr(\mathbf{x}|\mathbf{D}) = \frac{1}{Z}\exp(-\sum_{c\in\mathcal{C}}\psi_c(\mathbf{x}_c)), \tag{1}$$

where Z is a normalizing constant known as the partition function, and $\mathcal{C}$ is the set of all cliques. The term $\psi_c(\mathbf{x}_c)$ is known as the potential function of the clique c where $\mathbf{x}_c = \{x_i, i \in c\}$. The corresponding Gibbs energy is given by

$$E(\mathbf{x}) = -\log\Pr(\mathbf{x}|\mathbf{D}) - \log Z = \sum_{c\in\mathcal{C}}\psi_c(\mathbf{x}_c) \tag{2}$$

The maximum a posterior (MAP) labelling $\mathbf{x_{map}}$ of the random field is defined as

$$\mathbf{x_{map}} = \arg\max_{\mathbf{x}\in\mathbf{L}}\Pr(\mathbf{x}|\mathbf{D}) = \arg\min_{\mathbf{x}\in\mathbf{L}}E(\mathbf{x}). \tag{3}$$

2.1. Submodular Energy Functions

Submodular set functions play an important role in energy minimization as they can be minimized in polynomial time [2, 9]. In this paper we will explain their properties in terms of functions of binary random variables which can be seen as set functions [12].

Definition 1. *A projection of a function $f : \mathcal{L}^n \rightarrow \mathbb{R}$ on s variables is a function $f^p : \mathcal{L}^s \rightarrow \mathbb{R}$ which is obtained by fixing the values of $n - s$ arguments of $f(\cdot)$. Here p refers to the set of variables whose values have been fixed.*

Example 1. *The function $f^p(x_2, \ldots, x_n) = f(0, x_2, \ldots, x_n)$ is a projection of the function $f(x_1, x_2, \ldots, x_n)$.*

Definition 2. *A function of one binary variable is always submodular. A function $f(x_1, x_2)$ of two binary variables $\{x_1, x_2\}$ is submodular if and only if:*

$$f(0,0) + f(1,1) \leq f(0,1) + f(1,0) \tag{4}$$

A function $f : \mathcal{L}^n \rightarrow R$ is submodular if and only if all its projections on 2 variables are submodular [2, 12].

Minimizing submodular functions using graph cuts Certain submodular functions can be minimized by solving an st-mincut problem [2]. Kolmogorov *et al.* [12] showed that all submodular functions of binary variables which can be written in terms of potential function of cliques of sizes 2 and 3 can be minimized in this manner. Freedman and Drineas [7] extended this result by characterizing the class of functions $\mathcal{F}^n$ involving higher order cliques defined on binary variables whose minimization can be translated to an st-mincut problem. The class of multi-label submodular functions which can be translated into an st-mincut problem has also been characterized independently by [6] and [8].

2.2. Metric and Semi-metric Potential functions

In this subsection we provide the constraints for pairwise potentials to define a metric or a semi-metric.

Definition 3. *A potential function $\psi_{ij}(a, b)$ for a pairwise clique of two random variables $\{x_i, x_j\}$ is said to be a* semi-metric *if it satisfies*

$$\psi_{ij}(a,b) = 0 \iff a = b \tag{5}$$
$$\psi_{ij}(a,b) = \psi_{ij}(b,a) \geq 0 \tag{6}$$

Definition 4. *The potential function is* metric *if in addition to the above mentioned constraints it also satisfies*

$$\psi_{ij}(a,d) \leq \psi_{ij}(a,b) + \psi_{ij}(b,d). \tag{7}$$

Example 2. *The function $\psi_{ij}(a, b) = |a - b|^2$ is a semi-metric but not a metric as it does not always satisfy condition (7).*

3. Move Making Algorithms

In this section we describe the move making algorithms of [4] for approximate energy minimization and explain the conditions under which they can be applied.

3.1. Minimizing $\mathcal{P}^2$ functions

Boykov *et al.* [4] addressed the problem of minimizing energy functions consisting of unary and pairwise cliques. These functions can be written as

$$E(\mathbf{x}) = \sum_{i\in\mathcal{V}}\psi_i(x_i) + \sum_{i\in\mathcal{V}, j\in\mathcal{N}_j}\psi_{ij}(x_i, x_j). \tag{8}$$

They proposed two move making algorithms called α-expansions and $\alpha\beta$-swaps for this problem. These algorithms work by starting from a initial labelling $\mathbf{x}$ and making a series of changes (moves) which lower the energy iteratively. Convergence is achieved when the energy cannot

be decreased further. At each step the move decreasing the energy of the labelling by the most amount is made. We will refer to such a move as *optimal*. Recently an alternative interpretation of the α-expansions was given in [13].

Boykov *et al.* [4] showed that the optimal moves for certain energy functions of the form (8) can be computed by solving an st-mincut problem. Specifically, they showed that if the pairwise potential functions ψ_{ij} define a metric then the energy function in equation (8) can be approximately minimized using α-expansion. Similarly if ψ_{ij} defines a semi-metric, it can be minimized using $\alpha\beta$-swap.

3.2. Binary Moves and Move Energies

The moves of both the α-expansion and $\alpha\beta$-swap algorithms can be represented as a vector of binary variables $\mathbf{t} = \{t_i, \forall i \in \mathcal{V}\}$. A *transformation* function $T(\mathbf{x}, \mathbf{t})$ takes the current labelling $\mathbf{x}$ and a move $\mathbf{t}$ and returns the new labelling $\hat{\mathbf{x}}$ which has been induced by the move. The energy of a move $\mathbf{t}$ (denoted by $E_m(\mathbf{t})$) is defined as the energy of the labelling $\hat{\mathbf{x}}$ it induces i.e. $E_m(\mathbf{t}) = E(T(\mathbf{x}, \mathbf{t}))$. The optimal move is defined as $\mathbf{t}^* = \arg\min_{\mathbf{t}} E(T(\mathbf{x}, \mathbf{t}))$.

As discussed in section 2.1, the optimal move $\mathbf{t}^*$ can be computed in polynomial time if the function $E_m(\mathbf{t})$ is submodular. From definition 2 this implies that all projections of $E_m(\mathbf{t})$ on two variables should be submodular i.e.

$$E^p_m(0,0) + E^p_m(1,1) \leq E^p_m(0,1) + E^p_m(1,0), \ \ \forall p \in \mathcal{V} \times \mathcal{V}. \quad (9)$$

3.3. The α-expansion algorithm

An α-expansion move allows any random variable to either retain its current label or take label 'α'. One iteration of the algorithm involves performing expansions for all α in $\mathcal{L}$ in some order successively. The transformation function $T_\alpha(.)$ for an α-expansion move transforms the label of a random variable X_i as

$$T_\alpha(x_i, t_i) = \begin{cases} x_i & \text{if} \quad t_i = 0 \\ \alpha & \text{if} \quad t_i = 1. \end{cases} \quad (10)$$

The optimal α-expansion move can be computed in polynomial time if the energy function $E_\alpha(\mathbf{t}) = E(T_\alpha(\mathbf{x}, \mathbf{t}))$ satisfies constraint (9). Substituting the value of E_α in (9) we get the constraint

$$E^p(\alpha, \alpha) + E^p(x_i, x_j) \leq E^p(x_i, \alpha) + E^p(\alpha, x_j), \ \ \forall p \in \mathcal{V} \times \mathcal{V}. \quad (11)$$

3.4. The $\alpha\beta$-swap algorithm

An $\alpha\beta$-swap move allows a random variable whose current label is α or β to either take label α or β. One iteration of the algorithm involves performing swap moves for all α, β in $\mathcal{L}$ in some order successively. The transformation function $T_{\alpha\beta}()$ for an $\alpha\beta$-swap transforms the label of a random variable x_i as

$$T_{\alpha\beta}(x_i, t_i) = \begin{cases} \alpha & \text{if} \quad x_i = \alpha, \beta \text{ and } t_i = 0, \\ \beta & \text{if} \quad x_i = \alpha, \beta \text{ and } t_i = 1. \end{cases} \quad (12)$$

The optimal $\alpha\beta$-swap move can be computed in polynomial time if the energy function $E_{\alpha\beta}(\mathbf{t}) = E(T_{\alpha\beta}(\mathbf{x}, \mathbf{t}))$ satisfies (9). As before, substituting the value of $E_{\alpha\beta}$ in (9) we get the constraint

$$E^p(\alpha, \alpha) + E^p(\beta, \beta) \leq E^p(\beta, \alpha) + E^p(\alpha, \beta), \ \ \forall p \in \mathcal{V} \times \mathcal{V}. \quad (13)$$

In the next section we provide a class of energy functions which satisfy equation (11) and (13).

4. Characterizing $\mathcal{P}^n$ Functions

Now we characterize a class of higher order clique potentials for which the expansion and swap moves can be computed in polynomial time. Recall that $\mathcal{P}^n$ functions are defined on cliques of size at most n. From the additivity theorem [12] it follows that the optimal moves for all energy functions composed of these clique potentials can be computed in polynomial time. We constrain the clique potentials to take the form:

$$\psi_c(\mathbf{x}_c) = f_c(\mathcal{Q}_c(\oplus, \mathbf{x}_c)). \quad (14)$$

where $\mathcal{Q}_c(\oplus, \mathbf{x}_c)$ is a functional defined as:

$$\mathcal{Q}_c(\oplus, \mathbf{x}_c) = \oplus_{i,j \in c} \phi_c(x_i, x_j). \quad (15)$$

Here f_c is an arbitrary function of $\mathcal{Q}_c$, ϕ_c is a pairwise function defined on all pairs of random variables in the clique c, and $\oplus$ is an operator applied on these functions $\phi_c(x_i, x_j)$.

4.1. Conditions for $\alpha\beta$-swaps

We will now specify the constraints under which all $\alpha\beta$-swap moves for higher order clique potentials can be computed in polynomial time. For the moment we consider the case $\oplus = \sum$, i.e.

$$\mathcal{Q}_c(\mathbf{x}_c) = \sum_{i,j \in c} \phi_c(x_i, x_j). \quad (16)$$

Theorem 1. *The optimal $\alpha\beta$-swap move for any $\alpha, \beta \in \mathcal{L}$ can be computed in polynomial time if the potential function $\psi_c(\mathbf{x}_c)$ defined on the clique c is of the form (14) where $f_c(\cdot)$ is a concave[1] non-decreasing function, $\oplus = \sum$ and $\phi_c(\cdot, \cdot)$ satisfies the constraints*

$$\phi_c(a, b) = \phi_c(b, a) \quad \forall a, b \in \mathcal{L} \quad (17)$$

$$\phi_c(a, b) \geq \phi_c(d, d) \quad \forall a, b, d \in \mathcal{L} \quad (18)$$

Proof. To prove that the optimal swap move can be computed in polynomial time we need to show that all projections on two variables of any $\alpha\beta$-swap move energy are submodular. From equation (13) this implies that $\forall i, j \in c$ the condition:

$$\psi_c(\{\alpha, \alpha\} \cup \mathbf{x}_{c \setminus \{i,j\}}) + \psi_c(\{\beta, \beta\} \cup \mathbf{x}_{c \setminus \{i,j\}}) \leq \psi_c(\{\alpha, \beta\} \cup \mathbf{x}_{c \setminus \{i,j\}}) + \psi_c(\{\beta, \alpha\} \cup \mathbf{x}_{c \setminus \{i,j\}}) \quad (19)$$

[1] A function $f(x)$ is concave if for any two points (a, b) and λ where $0 \leq \lambda \leq 1$: $\lambda f(a) + (1 - \lambda) f(b) \leq f(\lambda a + (1 - \lambda) b)$.

should be satisfied. Here $\mathbf{x}_{c\setminus\{i,j\}}$ denotes the labelling of all variables $X_u, u \in c$ except i and j. The cost of any configuration $\{\alpha,\alpha\} \cup \mathbf{x}_{c\setminus\{i,j\}}$ of the clique can be written as

$$\begin{aligned} & \psi_c(\{x_i, x_j\} \cup \mathbf{x}_{c\setminus\{i,j\}}) = f_c(\mathcal{Q}_c(\{x_i, x_j\} \cup \mathbf{x}_{c\setminus\{i,j\}})) \\ = \; & f_c(\phi_c(x_i, x_j) + \mathcal{Q}_{c\setminus\{i,j\}}(\mathbf{x}_{c\setminus\{i,j\}}) + \\ & \sum_{u\in c\setminus\{i,j\}} \phi_c(x_i, \mathbf{x}_u) + \sum_{u\in c\setminus\{i,j\}} \phi_c(x_j, \mathbf{x}_u)) \end{aligned} \quad (20)$$

Let D represent $\mathcal{Q}_{c\setminus\{i,j\}}(\mathbf{x}_{c\setminus\{i,j\}})$, D_α represent $\sum_{u\in c\setminus\{i,j\}} \phi_c(\alpha, \mathbf{x}_u)$, and D_β represent $\sum_{u\in c\setminus\{i,j\}} \phi_c(\beta, \mathbf{x}_u)$. Using equation (20), the equation (19) becomes

$$\begin{aligned} & f_c(\phi_c(\alpha,\beta) + D_\alpha + D_\beta + D) \\ + \; & f_c(\phi_c(\beta,\alpha) + D_\beta + D_\alpha + D) \\ \geq \; & f_c(\phi_c(\alpha,\alpha) + 2D_\alpha + D) + f_c(\phi_c(\beta,\beta) + 2D_\beta + D). \end{aligned} \quad (21)$$

As $\phi_c(\beta,\alpha) = \phi_c(\alpha,\beta)$ from constraint (17) this condition transforms to:

$$\begin{aligned} & 2f_c(\phi_c(\alpha,\beta) + D_\alpha + D_\beta + D) \geq \\ & f_c(\phi_c(\alpha,\alpha) + 2D_\alpha + D) + f_c(\phi_c(\beta,\beta) + 2D_\beta + D). \end{aligned} \quad (22)$$

To prove (22) we need lemma 1.

Lemma 1. *For a non decreasing concave function $f(x)$:*

$$2c \geq a + b \implies 2f(c) \geq f(a) + f(b). \quad (23)$$

Proof in [10].

Using the above lemma together with the fact that

$$2\phi_c(\alpha,\beta) \geq \phi_c(\alpha,\alpha) + \phi_c(\beta,\beta) \quad \forall \alpha,\beta \in \mathcal{L} \quad (24)$$

(see constraint (18)), we see that the theorem hold true. □

The class of clique potentials described by theorem 1 is a strict generalization of the class specified by the constraints of [4] which can be obtained by considering only pairwise cliques, choosing $f_c()$ as a linear increasing function[2], and constraining $\phi_c(a,a) = 0, \forall a \in \mathcal{L}$.

4.2. Conditions for α-expansions

In this subsection we characterize the higher order clique potentials for which the optimal α-expansion move can be computed in polynomial time for all $\alpha \in \mathcal{L}, \mathbf{x} \in \mathbf{L}$.

Theorem 2. *The optimal α-expansion move for any $\alpha \in \mathcal{L}$ can be computed in polynomial time if the potential function $\psi_c(\mathbf{x}_c)$ defined on the clique c is of the form (14) where $f_c(\cdot)$ is a increasing linear function, $\oplus = \sum$ and $\phi_c(\cdot,\cdot)$ is a metric.*

[2] All linear function are concave.

Proof. To prove that the optimal expansion move can be computed in polynomial time we need to show that all projections of any α-expansion move energy on two variables of the clique are submodular. From equation (11) this implies that $\forall i, j \in c$ the condition

$$\begin{aligned} \psi_c(\{\alpha,\alpha\} \cup \mathbf{x}_{c\setminus\{i,j\}}) \; + \; & \psi_c(\{a,b\} \cup \mathbf{x}_{c\setminus\{i,j\}}) \leq \\ \psi_c(\{a,\alpha\} \cup \mathbf{x}_{c\setminus\{i,j\}}) \; + \; & \psi_c(\{\alpha,b\} \cup \mathbf{x}_{c\setminus\{i,j\}}) \end{aligned} \quad (25)$$

is satisfied. Here a and b are the current labels of the variables X_i and X_j respectively.

Let D represent $\mathcal{Q}_{c\setminus\{i,j\}}(\mathbf{x}_{c\setminus\{i,j\}})$, and D_l represent $\sum_{u\in c\setminus\{i,j\}} \phi_c(l, \mathbf{x}_u)$ for any label l. Then, using equation (20) the constraint (25) becomes

$$\begin{aligned} & f_c(\phi_c(\alpha,b) + D_\alpha + D_b + D) \\ + \; & f_c(\phi_c(a,\alpha) + D_a + D_\alpha + D) \\ \geq \; & f_c(\phi_c(\alpha,\alpha) + 2D_\alpha + D) \\ + \; & f_c(\phi_c(a,b) + D_a + D_b + D). \end{aligned} \quad (26)$$

Let $R_1 = \phi_c(\alpha,b) + D_\alpha + D_b + D$, $R_2 = \phi_c(a,\alpha) + D_a + D_\alpha + D$, $R_3 = \phi_c(\alpha,\alpha) + 2D_\alpha + D$, and $R_4 = f_c(\phi_c(a,b) + D_a + D_b + D)$. Since $\phi_c(\cdot,\cdot)$ is a metric, we observe that

$$\phi_c(\alpha,b) + \phi_c(a,\alpha) \geq \phi_c(\alpha,\alpha) + \phi_c(a,b) \quad (27)$$

$$\Rightarrow R_1 + R_2 \geq R_3 + R_4. \quad (28)$$

Thus, we require a function f such that

$$R_1 + R_2 \geq R_3 + R_4 \implies f(R_1) + f(R_2) \geq f(R_3) + f(R_4). \quad (29)$$

The following lemma provides us the form of this function.

Lemma 2. *For a function f, $y_1 + y_2 \geq y_3 + y_4 \implies f(y_1) + f(y_2) \geq f(y_3) + f(y_4)$ if and only if f is linear. Proof in [10].*

Since $f(\cdot)$ is linear, this proves the theorem. □

It should be noted that the class of clique potentials defined by the above theorem is a small subset of the class of functions which can be used under $\alpha\beta$-swaps. In fact it is the same class of energy function as defined by [4] i.e. $\mathcal{P}^2$. This can be seen by observing that the potentials of the higher order cliques defined by theorem 2 can be represented as a sum of metric pairwise clique potentials. This raises the question whether we can define a class of higher order clique potentials which cannot be decomposed into a set of $\mathcal{P}^2$ potentials and be solved using α-expansions. To answer this we define the $\mathcal{P}^n$ Potts model.

4.2.1 $\mathcal{P}^n$ Potts Model

We now introduce the $\mathcal{P}^n$ Potts model family of higher order clique potentials. This family is a strict generalization

of the Generalized Potts model [4] and can be used for modelling many problems in Computer Vision.

We define the $\mathcal{P}^n$ Potts model potential for cliques of size n as

$$\psi_c(\mathbf{x}_c) = \begin{cases} \gamma_k & \text{if} \quad x_i = l_k, \forall i \in c, \\ \gamma_{\max} & \text{otherwise.} \end{cases} \tag{30}$$

where $\gamma_{\max} > \gamma_k,\ \forall l_k \in \mathcal{L}$. For a pairwise clique this reduces to the $\mathcal{P}^2$ Potts model potential defined as $\psi_{ij}(a,b) = \gamma_k$ if $a = b = l_k$ and $\gamma_{\max}$ otherwise. If we use $\gamma_k = 0$, for all l_k, this function becomes an example of a *metric* potential function.

4.2.2 Going Beyond $\mathcal{P}^2$ for α-expansions

We now show how the class of potential functions characterized in section 4.2 can be extended by using: $\oplus =$'max' instead of $\oplus = \sum$ as in the previous subsection. To this end we define $\mathcal{Q}_c(\mathbf{x}_c)$ as

$$\mathcal{Q}_c(\mathbf{x}_c) = \max_{i,j \in c} \phi_c(x_i, x_j). \tag{31}$$

Theorem 3. *The optimal α-expansion move for any $\alpha \in \mathcal{L}$ can be computed in polynomial time if the potential function $\psi_c(\mathbf{x}_c)$ defined on the clique c is of the form (14) where $f_c(\cdot)$ is a increasing linear function, $\oplus$ = 'max' and $\phi_c(\cdot,\cdot)$ defines a $\mathcal{P}^2$ Potts Model.*

Proof. The cost of any configuration $\{\alpha, \alpha\} \cup \mathbf{x}_{c\setminus\{i,j\}}$ of the clique under $\oplus =$'max' can be written as

$$\begin{aligned} & \psi_c(\{x_i, x_j\} \cup \mathbf{x}_{c\setminus\{i,j\}}) \\ = \ & f_c(\mathcal{Q}_c(\{x_i, x_j\} \cup \mathbf{x}_{c\setminus\{i,j\}})) \qquad (32) \\ = \ & f_c(\max(\phi_c(x_i, x_j), \mathcal{Q}_{c\setminus\{i,j\}}(\mathbf{x}_{c\setminus\{i,j\}}), \\ & \max_{u \in c\setminus\{i,j\}} \phi_c(x_i, \mathbf{x}_u), \max_{u \in c\setminus\{i,j\}} \phi_c(x_j, \mathbf{x}_u))) \qquad (33) \end{aligned}$$

Substituting this value of ψ_c in constraint (25) and again using D to represent $\mathcal{Q}_{c\setminus\{i,j\}}(\mathbf{x}_{c\setminus\{i,j\}})$ and D_l represent $\sum_{u \in c\setminus\{i,j\}} \phi_c(l, \mathbf{x}_u)$ for any label l, we get:

$$\begin{aligned} & f_c(\max(\phi_c(\alpha, b), D_\alpha, D_b, D)) \\ + \ & f_c(\max(\phi_c(a, \alpha), D_a, D_\alpha, D)) \\ \geq \ & f_c(\max(\phi_c(\alpha, \alpha), D_\alpha, D_\alpha, D)) \\ + \ & f_c(\max(\phi_c(a, b), D_a, D_b, D)). \qquad (34) \end{aligned}$$

As f_c is a linear function, from lemma 2 we see that the above condition is true if:

$$\begin{aligned} & \max(\phi_c(\alpha, b), D_\alpha, D_b, D) + \max(\phi_c(a, \alpha), D_a, D_\alpha, D) \geq \\ & \max(\phi_c(\alpha, \alpha), D_\alpha, D_\alpha, D) + \max(\phi_c(a, b), D_a, D_b, D). \end{aligned}$$

We only consider the case $a \neq \alpha$ and $b \neq \alpha$. It can be easily seen that for all other cases the above inequality is satisfied by a equality. As ϕ_c is a $\mathcal{P}^2$ Potts model potential, the LHS of the above inequality is always equal to $2\gamma_{\max}$. As the maximum value of the RHS is $2\gamma_{\max}$ the above inequality is always true. □

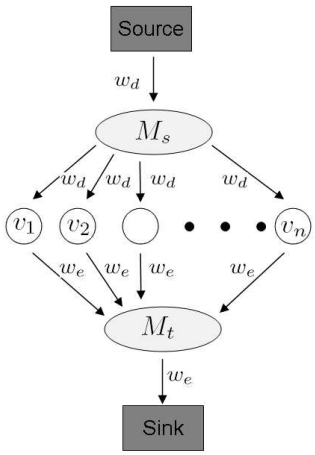


Figure 1. *Graph construction for computing the optimal moves for the $\mathcal{P}^n$ Potts model.*

Note that the class of potentials described in above theorem is the same as the family of clique potentials defined by the $\mathcal{P}^n$ Potts model in equation (30) for a clique c of size n. This proves that for the $\mathcal{P}^n$ Potts model the optimal α-expansion move can be solved in polynomial time. In fact we will show that the optimal α-expansion and $\alpha\beta$-swap moves for this subset of potential functions can be found by solving an st-mincut problem.

4.3. Graph Cuts for $\mathcal{P}^n$ Potts Model

We now consider the minimization of energy functions whose clique potentials form a $\mathcal{P}^n$ Potts model (see equation (30)). Specifically, we show that the optimal $\alpha\beta$-swap and α-expansion moves can be obtained by solving an st-mincut problem. The graph corresponding to the st-mincut is shown for only one clique. However, the additivity theorem [12] allows us to construct the graph for an arbitrary number of cliques by simply appending the graphs corresponding to single cliques.

$\alpha\beta$-swap : Given a clique $\mathbf{x}_c$, our aim is to find the optimal $\alpha\beta$-swap move (denoted by $\mathbf{t}_c^*$). Since the clique potential $\psi_c(\mathbf{x}_c)$ forms a $\mathcal{P}^n$ Potts model, its value after an $\alpha\beta$-swap move $\mathbf{t}_c = \{t_i, i \in c\}$ where $t_i \in \{0, 1\}$ is given by

$$\psi_c(T_{\alpha\beta}(\mathbf{x}_c, \mathbf{t}_c)) = \begin{cases} \gamma_\alpha & \text{if} \quad t_i = 0, \forall i \in c, \\ \gamma_\beta & \text{if} \quad t_i = 1, \forall i \in c, \\ \gamma_{\max} & \text{otherwise.} \end{cases} \tag{35}$$

Further, we can add a constant κ to all possible values of the clique potential without changing the optimal move $\mathbf{t}_c^*$. We choose $\kappa = \gamma_{\max} - \gamma_\alpha - \gamma_\beta$. Note that since $\gamma_{\max} \geq \gamma_\alpha$ and $\gamma_{\max} \geq \gamma_\beta$, the following hold true:

$$\gamma_\alpha + \kappa \geq 0, \quad \gamma_\beta + \kappa \geq 0, \tag{36}$$

$$\gamma_\alpha + \kappa + \gamma_\beta + \kappa = \gamma_{\max} + \kappa. \tag{37}$$

Without loss of generality, we assume $\mathbf{t}_c = \{t_1, t_2, \ldots, t_n\}$. Fig. 1 shows the graph construction corresponding to the above values of the clique potential. Here, the node v_i corresponds to t_i. In other words, after the computation of the st-mincut v_i is connected to the source (i.e. it belongs to the source set) if $t_i = 0$ and v_i is connected to the sink (i.e. it belongs to the sink set) if $t_i = 1$. In addition, there are two extra nodes denoted by M_s and M_t respectively. The weights of the graph are given by $w_d = \gamma_\beta + \kappa$ and $w_e = \gamma_\alpha + \kappa$. Note that all the weights are positive (see equations (36)). In order to show that this graph corresponds to the clique potential in equation (35) (plus the constant κ) we consider three cases:

- $t_i = 0, \forall i \in c$: In this case, the st-mincut corresponds to the edge connecting M_t with the sink which has a cost $w_e = \gamma_\alpha + \kappa$. Recall that the cost of an st-mincut is the sum of weights of the edges included in the st-mincut which go from the source set to the sink set.
- $t_i = 1, \forall i \in c$: In this case, the st-mincut corresponds to the edge connecting the source with M_s which has a cost $w_d = \gamma_\beta + \kappa$.
- All other cases: The st-mincut is given by the edges connecting M_t with the sink and the source with M_s. The cost of the cut is $w_d + w_e = \gamma_\alpha + \kappa + \gamma_\beta + \kappa = \gamma_{\max} + \kappa$ (from equation (37)).

Thus, we can find the optimal $\alpha\beta$-swap move for minimizing energy functions whose clique potentials form an $\mathcal{P}^n$ Potts model using an st-mincut operation.

α-expansion : Given a clique $\mathbf{x}_c$, our aim is to find the optimal α-expansion move $\mathbf{t}_c^*$. Again, since the clique potential $\psi_c(\mathbf{x}_c)$ forms an $\mathcal{P}^n$ Potts model, its value after an α-expansion move $\mathbf{t}_c$ is given by

$$\psi_c(T_\alpha(\mathbf{x}_c, \mathbf{t}_c)) = \begin{cases} \gamma_\alpha & \text{if} \quad t_i = 0, \forall i \in c, \\ \gamma & \text{if} \quad t_i = 1, \forall i \in c, \\ \gamma_{\max} & \text{otherwise,} \end{cases} \tag{38}$$

where $\gamma = \gamma_\beta$ if $x_i = \beta$ for all $i \in c$ and $\gamma = \gamma_{\max}$ otherwise. The above clique potential is similar to the one defined for the $\alpha\beta$-swap move in equation (35). Therefore, it can be represented using a graph by adding a constant $\kappa = \gamma_{\max} - \gamma_\alpha - \gamma$. This proves that the optimal α-expansion move can be obtained using an st-mincut operation.

5. Texture Based Segmentation

We now present experimental results which illustrates the advantage of higher order cliques. Higher order cliques provide a probabilistic formulation for a wide variety of exemplar based applications in computer vision, e.g. 3D reconstruction [19] and object recognition [14]. For this paper, we consider one such problem i.e. texture based segmentation.[3] This problem can be stated as follows. Given a set of distinct textures (e.g. a dictionary of RGB patches or histograms of textons) together with their object class labels, the task is to segment an image. In other words, the pixels of the image should be labelled as belonging to one of the object classes (e.g. see Fig. 3).

The above problem can be formulated within a probabilistic framework using CRF [15]. A CRF represents the conditional distribution of a set of random variables $\mathbf{X} = \{X_1, X_2, \ldots, X_N\}$ given the data $\mathbf{D}$. Each of the variables can take one label $x_i \in \{1, 2, \ldots, n_s\}$. In our case, n_s is the number of distinct object classes, a variable X_i represents a pixel $\mathbf{D}_i$ and $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ describes a segmentation. The most probable (i.e. maximum a posterior) segmentation can be obtained by (approximately) minimizing the corresponding Gibbs energy.

Pairwise CRF : For the problem of segmentation, it is common practice to assume a pairwise CRF where the cliques are of size at most two [1, 3, 22]. In this case, the Gibbs energy of the CRF is of the form:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j), \tag{39}$$

where $\mathcal{N}_i$ is the neighbourhood of pixel $\mathbf{D}_i$ (defined in this work as the 8-neighbourhood). The unary potential $\psi_i(x_i)$ is specified by the RGB distributions $\mathcal{H}_a, a = 1, \ldots, n_s$ of the segments as

$$\psi_i(x_i) = -\log p(\mathbf{D}_i | \mathcal{H}_a), \text{ when } x_i = a. \tag{40}$$

The pairwise potentials $\psi_{ij}(x_i, x_j)$ are defined such that they encourage contiguous segments whose boundaries lie on image edges, i.e.

$$\psi_{ij}(x_i, x_j) = \begin{cases} \lambda_1 + \lambda_2 \exp\left(\frac{-g^2(i,j)}{2\sigma^2}\right) & \text{if} \quad x_i \neq x_j, \\ 0 & \text{if} \quad x_i = x_j, \end{cases} \tag{41}$$

where λ_1, λ_2 and σ are some parameters. The term $g(i, j)$ represents the difference between the RGB values of pixels $\mathbf{D}_i$ and $\mathbf{D}_j$. We refer the reader to [3] for details. Note that the pairwise potentials $\psi_{ij}(x_i, x_j)$ form a metric. Hence, the energy function in equation (39) can be minimized using both $\alpha\beta$-swap and α-expansion algorithms.

Higher Order Cliques : The $\mathcal{P}^n$ functions presented in this paper allow us to go beyond the pairwise CRF framework by incorporating texture information as higher order cliques. Unlike the distributions $\mathcal{H}_a$ which describe the potential for one variable X_i, texture captures rich statistics of natural images [17, 24]. In this work, we represent the texture of each object class $s \in \{1, 2, \cdots, n_s\}$ using a dictionary $\mathbf{P}_s$ of $n_p \times n_p$ RGB patches. Note, however, that our

[3] Our forthcoming work also demonstrates the effectiveness of $\mathcal{P}^n$ functions on other applications.



Figure 2. *Segmented keyframe of the garden sequence. The left image shows the keyframe while the right image shows the corresponding segmentation provided by the user. The four different colours indicate pixels belonging to the four segments namely sky, house, garden and tree.*

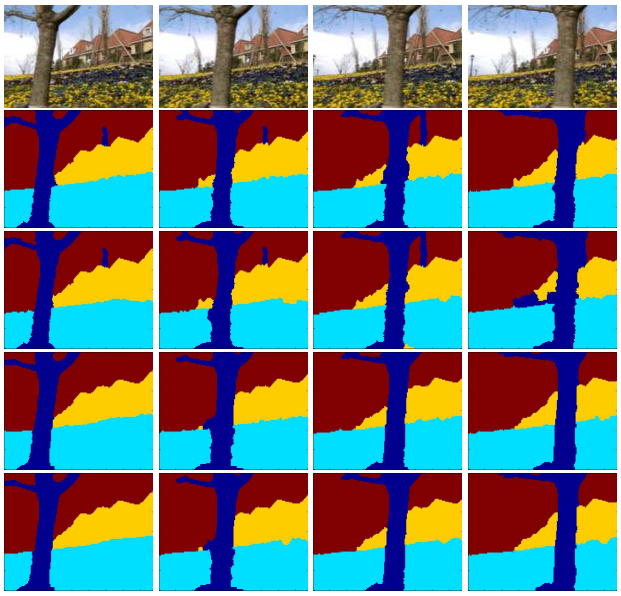


Figure 3. *The first row shows four frames of the garden sequence. The second row shows the segmentation obtained by minimizing the energy of the pairwise* CRF *(in equation (39)) using the $\alpha\beta$-swap algorithm. The four different colours indicate the four segments. The segmentations obtained using α-expansion to minimize the same energy are shown in the third row. The fourth row shows the results obtained by minimizing the energy containing higher order clique terms which form a $\mathcal{P}^n$ Potts model (given in equation (42)) using the $\alpha\beta$-swap algorithm. The fifth row shows the results obtained using the α-expansion algorithm to minimize the energy in equation (42). The use of higher order cliques results in more accurate segmentation.*

framework is independent of the representation of texture. As we will describe later, the likelihood of a patch of the image $\mathbf{D}$ belonging to the segment s can be computed using the dictionary $\mathbf{P}_s$.

The resulting texture based segmentation problem can be formulated using a CRF composed of higher order cliques. We define the Gibbs energy of this CRF as

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j) + \sum_{c \in \mathcal{C}} \psi_c(\mathbf{x}_c), \quad (42)$$

where c is a clique which represents the patch $\mathbf{D}_c = \{\mathbf{D}_i, i \in c\}$ of the image $\mathbf{D}$ and $\mathcal{C}$ is the set of all cliques. Note that we use overlapping patches $\mathbf{D}_c$ such that $|\mathcal{C}| = N$. The unary potentials $\psi_i(x_i)$ and the pairwise potentials $\psi_{ij}(x_i, x_j)$ are given by equations (40) and (41) respectively. The clique potentials $\psi_c(\mathbf{x}_c)$ are defined such that they form a $\mathcal{P}^n$ Potts model ($n = n_p^2$), i.e.

$$\psi_c(\mathbf{x}_c) = \begin{cases} \lambda_3 G(c, s) & \text{if } x_i = s, \forall i \in c, \\ \lambda_4 & \text{otherwise.} \end{cases} \quad (43)$$

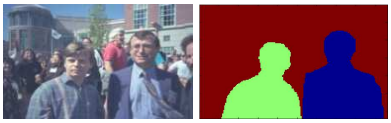


Figure 4. *The keyframe of the 'Dayton' video sequence segmented into three segments.*



Figure 5. *Segmentation results of the 'Dayton' sequence. Rows 2 and 3 show the results obtained for the frames shown in row 1 by minimizing the energy function in equation (39) using $\alpha\beta$-swap and α-expansion respectively. Row 4 and 5 show the segmentations obtained by minimizing the energy in equation (42) using $\alpha\beta$-swap and α-expansion respectively. The use of higher order cliques results in more accurate segmentation.*

Here $G(c, s)$ is the minimum difference between the RGB values of patch $\mathbf{D}_c$ and all patches belonging to the dictionary $\mathbf{P}_s$. Note that the above energy function encourages the patch $\mathbf{D}_c$ which are similar to a patch in $\mathbf{P}_s$ to take the label s. Since the clique potentials form a $\mathcal{P}^n$ Potts model, they can be minimized using the $\alpha\beta$-swap and α-expansion algorithms as described in section 4.3.

Results : We tested our approach for segmenting frames of a video sequence. A *keyframe* of the video was manually segmented and used to learn the distributions $\mathcal{H}_a$ and the dictionary of patches $\mathbf{P}_s$. The $\alpha\beta$-swap and α-expansion algorithms were used to perform segmentation on the other frames. In all our experiments, we used patches of size 4×4, together with the following parameter setting: $\lambda_1 = 0.6$, $\lambda_2 = 6$, $\lambda_3 = 0.6$, $\lambda_4 = 6.5$ and $\sigma = 5$.

Fig. 2 shows the segmented keyframe of the well-known garden sequence. Fig. 3 (row 2) shows the segmentation obtained for four frames by minimizing the energy function of the pairwise CRF (defined in equation (39)) using the $\alpha\beta$-swap algorithm. Note that these frames are different from the keyframe (see Fig. 3 (row 1)). The results obtained by the α-expansion algorithm are shown in Fig. 3 (row 3). The α-expansion algorithm takes an average of 3.7 seconds per frame compared to the 4.7 seconds required by the $\alpha\beta$-swap

algorithm. Note that the segmentations obtained by both the algorithms are inaccurate due to small clique sizes.

Fig. 3 (row 4) shows the segmentations obtained when the energy function of the higher order CRF (defined in equation (42)) is minimized using $\alpha\beta$-swap. Fig. 3 (row 5) shows the results obtained using the α-expansion algorithm. On average, α-expansion takes 4.42 seconds while $\alpha\beta$-swap takes 5 seconds which is comparable to the case when the pairwise CRF is used. For both $\alpha\beta$-swap and α-expansion, the use of higher order cliques provides more accurate segmentation than the pairwise CRF formulation.

Fig. 4 shows another example of a segmented keyframe from a video sequence. The segmentations obtained for four frames of this video are shown in Fig. 5. Note that even though we do not use motion information, the segmentations provided by higher order cliques are comparable to the methods based on layered motion segmentation.

6. Discussion and Conclusions

In this paper we have characterized a class of higher order clique potentials for which the optimal expansion and swap moves can be computed in polynomial time. We also introduced the $\mathcal{P}^n$ Potts model family of clique potentials and showed that the optimal moves for it can be solved using graph cuts. Their use is demonstrated on the texture based video segmentation problem. The $\mathcal{P}^n$ Potts model potentials can be used to solve many other Computer Vision problems such as object recognition [14] and novel view synthesis [5] as will be shown in forthcoming works.

We conclude with the observation that the optimal moves for many interesting clique potentials such as those that preserve planarity are NP-hard to compute [10]. Hence, they do not lend themselves to efficient move making algorithms.

Acknowledgements This work was supported by EPSRC research grants GR/T21790/01(P) and EP/C006631/1(P) and the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors' views.

References

[1] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. In *ECCV*, pages I: 428–441, 2004. 6

[2] E. Boros and P. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155–225, 2002. 2

[3] Y. Boykov and M. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *ICCV*, pages I: 105–112, 2001. 6

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001. 1, 2, 3, 4, 5

[5] A. W. Fitzgibbon, Y. Wexler, and A. Zisserman. Image-based rendering using image-based priors. In *ICCV*, pages 1176–1183, 2003. 8

[6] B. Flach. Strukturelle bilderkennung. Technical report, Universit at Dresden, 2002. 2

[7] D. Freedman and P. Drineas. Energy minimization via graph cuts: Settling what is possible. In *CVPR*, 2005. 1, 2

[8] H. Ishikawa. Exact optimization for markov random fields with convex priors. *PAMI*, 25:1333–1336, October 2003. 2

[9] S. Iwata, S. T. McCormick, and M. Shigeno. A strongly polynomial cut canceling algorithm for the submodular flow problem. *Lecture Notes in Computer Science*, 1610, 1999. 2

[10] P. Kohli, M. P. Kumar, and P. H. S. Torr. Solving energies with higher order cliques. Technical report, Oxford Brookes University, 2007. 4, 8

[11] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10), 2006. 1

[12] V. Kolmogorov and R. Zabih. What energy functions can be minimizedvia graph cuts?. *PAMI*, 26(2):147–159, 2004. 1, 2, 3, 5

[13] N. Komodakis and G. Tziritas. A new framework for approximate labeling via graph cuts. In *ICCV*, 2005. 3

[14] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Extending pictorial structures for object recognition. In *BMVC*, pages II: 789–798, 2004. 6, 8

[15] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labelling sequence data. In *ICML*, 2001. 1, 6

[16] X. Lan, S. Roth, D. P. Huttenlocher, and M. J. Black. Efficient belief propagation with learned higher-order markov random fields. In *ECCV (2)*, pages 269–282, 2006. 1

[17] T. Leung and J. Malik. Recognizing surfaces using three-dimensional textons. In *ICCV*, pages 1010–1017, 1999. 6

[18] T. Meltzer, C. Yanover, and Y. Weiss. Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. In *ICCV*, pages 428–435, 2005. 1

[19] A. Neubeck, A. Zalesny, and L. van Gool. 3d texture reconstruction from extensive BTF data. In *Texture*, pages 13–18, 2005. 6

[20] R. Paget and I. D. Longstaff. Texture synthesis via a noncausal nonparametric multiscale markov random field. *IEEE Transactions on Image Processing*, 7(6):925–931, 1998. 1

[21] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *CVPR*, pages 860–867, 2005. 1

[22] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: interactive foreground extraction using iterated graph cuts. In *SIGGRAPH*, pages 309–314, 2004. 6

[23] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. F. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields. In *ECCV (2)*, pages 16–29, 2006. 1

[24] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *CVPR*, pages 691–698, 2003. 6

[25] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Tree-based reparameterization for approximate inference on loopy graphs. In *NIPS*, pages 1001–1008, 2001. 1

[26] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propagation. In *NIPS*, pages 689–695, 2000. 1