

# $\mathcal{P}^3$ & Beyond: Move Making Algorithms for Solving Higher Order Functions

Pushmeet Kohli, *Member, IEEE*, M. Pawan Kumar, *Member, IEEE*,  
Philip H. S. Torr, *Senior Member, IEEE*,

## Abstract

In this paper we extend the class of energy functions for which the optimal  $\alpha$ -expansion and  $\alpha\beta$ -swap moves can be computed in polynomial time. Specifically, we introduce a novel family 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  $\alpha$ -expansion and  $\alpha\beta$ -swap move making algorithms 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 image or video segmentation problem.

## I. 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 [5], [12], tree-reweighted message passing [11], [29] and variants of belief propagation (BP) [18], [31]. These algorithms allow us to perform approximate inference on graphical models such as Markov Random Fields (MRF) and Conditional Random Fields (CRF) [15].

Two of the most popular *move* making algorithms for approximate energy minimization are  $\alpha$ -expansion and  $\alpha\beta$ -swap [5]. They are extremely efficient and have been shown to produce good results for a number of problems [27]. 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 an *optimal* move (i.e. a 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.* [5] 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 potential functions defined over cliques of size at most two. We call this class of energy functions  $\mathcal{P}^2$ . In this paper we extend the results of [5] by providing a characterization of energy functions

involving higher order cliques, i.e. cliques of size 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  $n$  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 [6], [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 the MRF and CRF frameworks [16], [20], [22]. 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 experiments indicate that BP gave results comparable to standard gradient descent. In contrast, we use powerful move making algorithms such as  $\alpha$ -expansion and  $\alpha\beta$ -swaps to minimize such functions. In this paper, we provide a characterization of energy functions defined on cliques of size 3 or more which can be solved using these algorithms. More specifically, we prove that the optimal  $\alpha$ -expansion and  $\alpha\beta$ -swap moves for this class of functions can be computed in polynomial time.

It should be noted that our results are a generalization of the class of energy functions specified by [5]. We also give examples of higher order potential functions for which it is NP-hard to compute the optimal move. The previous version of this paper appeared as [9]. This extended version contains a discussion of higher order potentials for which it is NP-hard to compute the optimal moves. It also contains proofs of lemmas which were missing in the earlier version.

**Organization of the Paper:** In section II, we provide the notation and discuss the basic theory of energy minimization and submodular functions. Section III describes the  $\alpha$ -expansion and  $\alpha\beta$ -swap algorithms. Further, it provides constraints on the pairwise potentials which guar-

antee computation of the optimal move in polynomial time [5]. In section IV, 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 V). In section VI, we give an example of a useful higher order potential for which the move energies are non-submodular and thus the problem of computing the optimal move is NP-hard in general. We conclude in section VII by listing some Computer Vision problems where higher order clique potentials can be used.

## II. PRELIMINARIES

Many problems in computer vision and artificial intelligence can be formulated in terms of minimizing an energy function  $E(\mathbf{x})$ . This function is defined on a set of discrete random variables  $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$  each of which takes a value from the label set  $\mathcal{L} = \{l_1, l_2, \dots, l_k\}$ . A labelling or *configuration*  $\mathbf{x} \in \mathcal{L}^N$  is a possible assignment of labels to the random variables.

The energy function associated with a labelling problem is defined in terms of the posterior probability distribution of the possible configurations of the random variables, i.e.

$$E(\mathbf{x}) = -\log \Pr(\mathbf{x}|\mathbf{D}) - \log Z, \quad (1)$$

where  $Z$  is the partition function. Further, they can be written in terms of a sum of potential functions defined over cliques of variables as:

$$E(\mathbf{x}) = \sum_{c \in \mathcal{C}} \psi_c(\mathbf{x}_c), \quad (2)$$

where  $\psi_c(\mathbf{x}_c)$  is the potential function defined over the clique  $c$  where  $\mathbf{x}_c = \{x_i, i \in c\}$ . Given the above notation, the maximum a posterior (MAP) labelling  $\mathbf{x}_{\text{map}}$  of the random variables can be found as:

$$\mathbf{x}_{\text{map}} = \arg \max_{\mathbf{x} \in \mathcal{L}^N} \Pr(\mathbf{x}|\mathbf{D}) = \arg \min_{\mathbf{x} \in \mathcal{L}^N} E(\mathbf{x}). \quad (3)$$

### A. Submodular Energy Functions

Submodular set functions play an important role in energy minimization as they can be minimized in polynomial time [2], [8]. We will explain their properties in terms of functions of binary random variables which can be seen as set functions [12]. In order to describe submodular energy functions we require the following definition of a projection of a function.

*Projection:* 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  $f(\cdot)$ . For example, the function  $f^p(x_2, \dots, x_N) = f(0, x_2, \dots, x_N)$  is a projection of the function  $f(x_1, x_2, \dots, x_N)$ . Using the concept of projection, a submodular function is defined as follows.

*Submodularity:* 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:

$$f(0, 0) + f(1, 1) \leq f(0, 1) + f(1, 0) \quad (4)$$

A function  $f : \{0, 1\}^N \rightarrow R$  is submodular if and only if all its projections on 2 variables are submodular [2], [12].

### B. 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 [6] extended this result by characterizing a class of functions  $\mathcal{F}^n$  involving higher order cliques defined on binary variables whose minimization can be transformed 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 in [7], [21], [25].

### C. Metric and Semi-metric Potential Functions

We now provide the constraints under which pairwise potentials are said to define a metric or a semi-metric.

*Semimetric:* 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*  $\psi$  if for all  $a, b \in \mathcal{L}$  it satisfies

$$\psi_{ij}(a, b) = 0 \iff a = b \quad (\text{identity of indiscernibles}) \quad (5)$$

$$\psi_{ij}(a, b) \geq 0 \quad (\text{non-negativity}) \quad (6)$$

$$\psi_{ij}(a, b) = \psi_{ij}(b, a) \quad (\text{symmetry}) \quad (7)$$

*Metric:* The potential function is *metric* if in addition to the above mentioned constraints it also satisfies the triangle inequality

$$\psi_{ij}(a, d) \leq \psi_{ij}(a, b) + \psi_{ij}(b, d), \quad \forall a, b, d \in \mathcal{L}. \quad (8)$$

For example, the function  $\psi_{ij}(a, b) = |a - b|^2$  is a semi-metric but not a metric as it does not always satisfy the triangle inequality (8).

### III. MOVE MAKING ALGORITHMS

We now describe the move making algorithms of [5] for approximate energy minimization and explain the conditions under which they can be applied.

#### A. Minimizing $\mathcal{P}^2$ functions

Boykov *et al.* [5] addressed the problem of minimizing energy functions consisting of unary and pairwise cliques. These functions can be written as

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j), \quad (9)$$

where  $\mathcal{N}_i$  contains the indices of the set of neighbouring random variables of  $X_i$  (i.e.  $j \in \mathcal{N}_i$  if, and only if,  $X_i$  and  $X_j$  are neighbours). They proposed two move making algorithms called  $\alpha$ -expansion and  $\alpha\beta$ -swap for this problem. These algorithms work by starting from an 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 a move decreasing the energy of the labelling by the most amount is made. We will refer to such a move as *optimal*.

#### B. Binary Moves and Move Energies

The moves of both the  $\alpha$ -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.  $E_m(\mathbf{t}) = E(T(\mathbf{x}, \mathbf{t}))$ . An optimal move is then defined as  $\mathbf{t}^* = \arg \min_{\mathbf{t}} E(T(\mathbf{x}, \mathbf{t}))$ .

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

$$E_m^p(0, 0) + E_m^p(1, 1) \leq E_m^p(0, 1) + E_m^p(1, 0), \quad \forall p \in \mathcal{V} \times \mathcal{V}. \quad (10)$$

### C. The $\alpha$ -expansion algorithm

An  $\alpha$ -expansion move allows any random variable to either retain its current label or take label ' $\alpha$ '. One iteration of the algorithm involves performing expansions for all  $\alpha$  in  $\mathcal{L}$  in some order successively. Recently an alternative interpretation of  $\alpha$ -expansions was given in [13].

The transformation function  $T_\alpha(\cdot)$  for an  $\alpha$ -expansion move transforms the label of a random variable  $X_i$  as

$$T_\alpha(x_i, t_i) = \begin{cases} x_i & \text{if } t_i = 0 \\ \alpha & \text{if } t_i = 1. \end{cases} \quad (11)$$

The optimal  $\alpha$ -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 (10). Substituting the value of  $E_\alpha$  in (10) we get the constraint

$$E^p(x_i, x_j) + E^p(\alpha, \alpha) \leq E^p(x_i, \alpha) + E^p(\alpha, x_j), \quad \forall p \in \mathcal{V} \times \mathcal{V}. \quad (12)$$

Boykov *et al.* [5] showed that if the pairwise potential functions  $\psi_{ij}$  define a metric then the energy function in equation (9) can be approximately minimized using the  $\alpha$ -expansion algorithm. This easily follows by observing that all metric potentials satisfy equation (12).

### D. The $\alpha\beta$ -swap algorithm

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

$$T_{\alpha\beta}(x_i, t_i) = \begin{cases} \alpha & \text{if } x_i = \alpha \text{ or } \beta \text{ and } t_i = 0, \\ \beta & \text{if } x_i = \alpha \text{ or } \beta \text{ and } t_i = 1. \end{cases} \quad (13)$$

The optimal  $\alpha\beta$ -swap move can be computed in polynomial time if the energy function

$$E_{\alpha\beta}(\mathbf{t}) = E(T_{\alpha\beta}(\mathbf{x}, \mathbf{t})) \quad (14)$$

satisfies (10). As before, substituting the value of  $E_{\alpha\beta}$  in (10) we get the constraint

$$E^p(\alpha, \alpha) + E^p(\beta, \beta) \leq E^p(\alpha, \beta) + E^p(\beta, \alpha), \quad \forall p \in \mathcal{V} \times \mathcal{V}. \quad (15)$$

Boykov *et al.* [5] showed that if the pairwise potential functions  $\psi_{ij}$  define a semimetric then the energy function in equation (9) can be approximately minimized using  $\alpha\beta$ -swaps. This result can be easily verified by observing that all semi-metric potentials satisfy equation (15).

In the next section we show how the above mentioned move algorithms can be used to minimize higher order energy functions.

#### IV. 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:

$$\psi_c(\mathbf{x}_c) = f_c(\mathcal{Q}_c(\oplus, \mathbf{x}_c)). \quad (16)$$

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

$$\mathcal{Q}_c(\oplus, \mathbf{x}_c) = \oplus_{i,j \in c} \phi_c(x_i, x_j). \quad (17)$$

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

##### A. 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.

$$\mathcal{Q}_c(\mathbf{x}_c) = \sum_{i,j \in c} \phi_c(x_i, x_j). \quad (18)$$

*Theorem 1:* The optimal  $\alpha\beta$ -swap move for any  $\alpha$  and  $\beta \in \mathcal{L}$  can be computed in polynomial time for a potential function  $\psi_c(\mathbf{x}_c)$  of the form (16) if  $f_c(\cdot)$  is a concave<sup>1</sup> non-decreasing function,  $\oplus = \sum$  and  $\phi_c(\cdot, \cdot)$  satisfies the constraints

$$\phi_c(a, b) = \phi_c(b, a) \quad \forall a, b \in \mathcal{L} \quad (19)$$

$$\phi_c(a, b) \geq \phi_c(d, d) \quad \forall a, b, d \in \mathcal{L} \quad (20)$$

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

*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 (15) this implies that for all  $i$  and  $j$  in  $c$  the condition:

$$\begin{aligned} \psi_c(\{\alpha, \alpha\} \cup \mathbf{x}_{c \setminus \{i, j\}}) + \psi_c(\{\beta, \beta\} \cup \mathbf{x}_{c \setminus \{i, j\}}) \leq \\ \psi_c(\{\alpha, \beta\} \cup \mathbf{x}_{c \setminus \{i, j\}}) + \psi_c(\{\beta, \alpha\} \cup \mathbf{x}_{c \setminus \{i, j\}}) \end{aligned} \quad (21)$$

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  $\{x_i, x_j\} \cup \mathbf{x}_{c \setminus \{i, j\}}$  of the clique can be written as

$$\begin{aligned} \psi_c(\{x_i, x_j\} \cup \mathbf{x}_{c \setminus \{i, j\}}) &= f_c(\mathcal{Q}_c(\{x_i, x_j\} \cup \mathbf{x}_{c \setminus \{i, j\}})) \\ &= f_c(\phi_c(x_i, x_j) + \mathcal{Q}_{c \setminus \{i, j\}}(\mathbf{x}_{c \setminus \{i, j\}}) + \\ &\quad \sum_{u \in c \setminus \{i, j\}} \phi_c(x_i, x_u) + \sum_{u \in c \setminus \{i, j\}} \phi_c(x_j, x_u)) \end{aligned} \quad (22)$$

Let  $D$  represent the term  $\mathcal{Q}_{c \setminus \{i, j\}}(\mathbf{x}_{c \setminus \{i, j\}})$ ,  $D_\alpha$  represent  $\sum_{u \in c \setminus \{i, j\}} \phi_c(\alpha, x_u)$ , and  $D_\beta$  represent  $\sum_{u \in c \setminus \{i, j\}} \phi_c(\beta, x_u)$ . Using equation (22), the equation (21) becomes

$$\begin{aligned} f_c(\phi_c(\alpha, \beta) + D_\alpha + D_\beta + D) + f_c(\phi_c(\beta, \alpha) + D_\beta + D_\alpha + D) \\ \geq f_c(\phi_c(\alpha, \alpha) + 2D_\alpha + D) + f_c(\phi_c(\beta, \beta) + 2D_\beta + D). \end{aligned} \quad (23)$$

As  $\phi_c(\beta, \alpha) = \phi_c(\alpha, \beta)$  from constraint (19) this condition transforms to:

$$\begin{aligned} 2f_c(\phi_c(\alpha, \beta) + D_\alpha + D_\beta + D) \geq \\ f_c(\phi_c(\alpha, \alpha) + 2D_\alpha + D) + f_c(\phi_c(\beta, \beta) + 2D_\beta + D). \end{aligned} \quad (24)$$

For a non decreasing concave function  $f(\cdot)$ ,

$$2c \geq a + b \implies 2f(c) \geq f(a) + f(b), \quad \forall a, b, c. \quad (25)$$

Using this property together with the fact that

$$2\phi_c(\alpha, \beta) \geq \phi_c(\alpha, \alpha) + \phi_c(\beta, \beta) \quad \forall \alpha, \beta \in \mathcal{L} \quad (26)$$

(see constraint (20)), we see that the theorem holds true.  $\blacksquare$

The class of clique potentials described by Theorem 1 is a strict generalization of the class specified by the constraints of [5]. This can be verified by considering only pairwise cliques, choosing  $f_c(\cdot)$  as a linear increasing function, and constraining  $\phi_c(a, a) = 0, \forall a \in \mathcal{L}$ . Recall that all linear functions are concave.

### B. Conditions for $\alpha$ -expansions

We now characterize the higher order clique potentials for which the optimal  $\alpha$ -expansion move can be computed in polynomial time for all  $\alpha \in \mathcal{L}, \mathbf{x} \in \mathcal{L}^N$ .

*Theorem 2:* Let the potential function  $\psi_c(\mathbf{x}_c)$  be of the form (16) such that  $\oplus = \sum$ ,  $\phi_c(\cdot, \cdot)$  is a metric and  $f_c(\cdot)$  is continuous. Then, the optimal  $\alpha$ -expansion move for any  $\alpha \in \mathcal{L}$  can be computed in polynomial time if, and only if,  $f_c(\cdot)$  is an increasing linear function.

*Proof:* To prove that the optimal expansion move can be computed in polynomial time we need to show that all projections of any  $\alpha$ -expansion move energy on two variables of the clique are submodular. From equation (12) this implies that  $\forall i, j \in c$  the condition

$$\begin{aligned} \psi_c(\{\alpha, \alpha\} \cup \mathbf{x}_{c \setminus \{i, j\}}) + \psi_c(\{a, b\} \cup \mathbf{x}_{c \setminus \{i, j\}}) &\leq \\ \psi_c(\{a, \alpha\} \cup \mathbf{x}_{c \setminus \{i, j\}}) + \psi_c(\{\alpha, b\} \cup \mathbf{x}_{c \setminus \{i, j\}}) &\end{aligned} \quad (27)$$

must be 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  $\max_{u \in c \setminus \{i, j\}} \phi_c(l, x_u)$  for any label  $l$ . Then, using equation (22) the constraint (27) becomes

$$\begin{aligned} &f_c(\phi_c(\alpha, b) + D_\alpha + D_b + D) \\ &+ f_c(\phi_c(a, \alpha) + D_a + D_\alpha + D) \\ &\geq f_c(\phi_c(\alpha, \alpha) + 2D_\alpha + D) \\ &+ f_c(\phi_c(a, b) + D_a + D_b + D). \end{aligned} \quad (28)$$

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  $R_4 = \phi_c(a, b) + D_a + D_b + D$ . Since  $\phi_c(\cdot, \cdot)$  is a metric, we observe that

$$\phi_c(\alpha, b) + \phi_c(a, \alpha) \geq \phi_c(\alpha, \alpha) + \phi_c(a, b) \quad (29)$$

$$\Rightarrow R_1 + R_2 \geq R_3 + R_4. \quad (30)$$

Thus, we require a function  $f$  such that

$$R_1 + R_2 \geq R_3 + R_4 \implies f(R_1) + f(R_2) \geq f(R_3) + f(R_4). \quad (31)$$

The following lemma provides us the form of this function.

*Lemma 1:* For a continuous function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$y_1 + y_2 \geq y_3 + y_4 \implies f(y_1) + f(y_2) \geq f(y_3) + f(y_4) \quad (32)$$

for all  $y_1, y_2, y_3, y_4 \in \mathbb{R}$  if, and only if,  $f(\cdot)$  is linear. Proof in Appendix.

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

It should be noted that the class of higher order potentials specified by the above theorem is a small subset of the family of functions which can be used under  $\alpha\beta$ -swap (characterized in Theorem 1). In fact it is the same class of energy functions which was specified in [5], i.e.  $\mathcal{P}^2$ . This can be verified by observing that the potentials specified by Theorem 2 can be represented as a sum of metric pairwise functions. 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 can still be solved using  $\alpha$ -expansion. To answer this we need to define the  $\mathcal{P}^n$  Potts model.

**$\mathcal{P}^n$  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 [5] 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

$$\psi_c(\mathbf{x}_c) = \begin{cases} \gamma_k & \text{if } x_i = l_k, \forall i \in c, \\ \gamma_{\max} & \text{otherwise.} \end{cases} \quad (33)$$

where  $\gamma_{\max} \geq \gamma_k, \forall l_k \in \mathcal{L}$ . 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. Further, if we use  $\gamma_k = 0$ , for all  $l_k$ , this function becomes an example of a *metric* potential function.

1) *Going Beyond  $\mathcal{P}^2$  for  $\alpha$ -expansions:* We now show how the class of potential functions characterized in section IV-B can be extended by using:  $\oplus = \text{'max'}$  instead of  $\oplus = \sum$  as in the previous subsection. To this end we define  $\mathcal{Q}_c(\mathbf{x}_c)$  as

$$\mathcal{Q}_c(\mathbf{x}_c) = \max_{i,j \in c} \phi_c(x_i, x_j). \quad (34)$$

*Theorem 3:* The optimal  $\alpha$ -expansion move for any  $\alpha \in \mathcal{L}$  can be computed in polynomial time for a potential function  $\psi_c(\mathbf{x}_c)$  of the form (16) if  $f_c(\cdot)$  is an increasing linear function,  $\oplus = \text{'max'}$  and  $\phi_c(\cdot, \cdot)$  defines a  $\mathcal{P}^2$  Potts Model.

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

$$\begin{aligned} \psi_c(\{x_i, x_j\} \cup \mathbf{x}_{c \setminus \{i,j\}}) &= f_c(\mathcal{Q}_c(\{x_i, x_j\} \cup \mathbf{x}_{c \setminus \{i,j\}})) \\ &= f_c(\max(\phi_c(x_i, x_j), \mathcal{Q}_{c \setminus \{i,j\}}(\mathbf{x}_{c \setminus \{i,j\}}), \\ &\quad \max_{u \in c \setminus \{i,j\}} \phi_c(x_i, x_u), \max_{u \in c \setminus \{i,j\}} \phi_c(x_j, x_u))) \end{aligned} \quad (35)$$

Substituting this value of  $\psi_c$  in constraint (27) and again using  $D$  to represent  $\mathcal{Q}_{c \setminus \{i,j\}}(\mathbf{x}_{c \setminus \{i,j\}})$  and  $D_l$  represent  $\sum_{u \in c \setminus \{i,j\}} \phi_c(l, x_u)$  for any label  $l$ , we get:

$$\begin{aligned} &f_c(\max(\phi_c(\alpha, b), D_\alpha, D_b, D)) \\ &+ f_c(\max(\phi_c(a, \alpha), D_a, D_\alpha, D)) \\ &\geq f_c(\max(\phi_c(\alpha, \alpha), D_\alpha, D_\alpha, D)) \\ &+ f_c(\max(\phi_c(a, b), D_a, D_b, D)). \end{aligned} \quad (36)$$

As  $f_c$  is a linear function, from lemma 1 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  $b \neq \alpha$ . For all other cases it can be easily seen that the above inequality is satisfied by an equality. As  $\phi_c$  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. This proves the Theorem.  $\blacksquare$

The class of higher order potentials specified by the above Theorem is the same as the family of clique potentials defined by the  $\mathcal{P}^n$  Potts model (33) for a clique  $c$  of size  $n$ . This proves that the optimal  $\alpha$ -expansion move for the  $\mathcal{P}^n$  Potts model can be computed in polynomial time. In the next section we will show how the optimal  $\alpha$ -expansion and  $\alpha\beta$ -swap moves for this subset of potential functions can be found by solving an st-mincut problem.

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

We now consider the minimization of energy functions whose clique potentials take the form a  $\mathcal{P}^n$  Potts model (see equation (33)). Specifically, we show that the optimal  $\alpha\beta$ -swap and  $\alpha$ -expansion moves for such potential functions can be computed by solving an st-mincut problem.

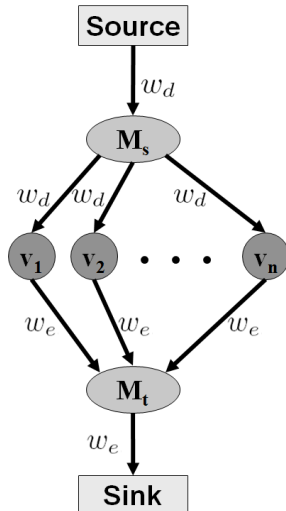


Fig. 1. Graph construction for computing the optimal moves for the  $\mathcal{P}^n$  Potts model.

The graph in which the st-mincut needs to be computed is shown for only a single clique potential. However, the additivity theorem [12] allows us to construct the graph for an arbitrary number of potentials by simply merging those corresponding to individual cliques.

**$\alpha\beta$ -swap:** Given a clique  $c$ , 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, we do not need to consider the move from a configuration in which any variable in the clique is assigned a label other than  $\alpha$  or  $\beta$ . In this scenario the clique potential only adds a constant to the  $\alpha\beta$ -swap move energy and thus can be ignored without changing the optimal move. For all other configurations, the potential function after an  $\alpha\beta$ -swap move  $\mathbf{t}_c = \{t_i, i \in c\}$  (where  $t_i \in \{0, 1\}$ ) is given by

$$\psi_c(T_{\alpha\beta}(\mathbf{x}_c, \mathbf{t}_c)) = \begin{cases} \gamma_\alpha & \text{if } t_i = 0, \forall i \in c, \\ \gamma_\beta & \text{if } t_i = 1, \forall i \in c, \\ \gamma_{\max} & \text{otherwise.} \end{cases} \quad (37)$$

Further, we can add a constant  $\kappa$  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:

$$\gamma_\alpha + \kappa \geq 0, \quad \gamma_\beta + \kappa \geq 0, \quad (38)$$

$$\gamma_\alpha + \kappa + \gamma_\beta + \kappa = \gamma_{\max} + \kappa. \quad (39)$$

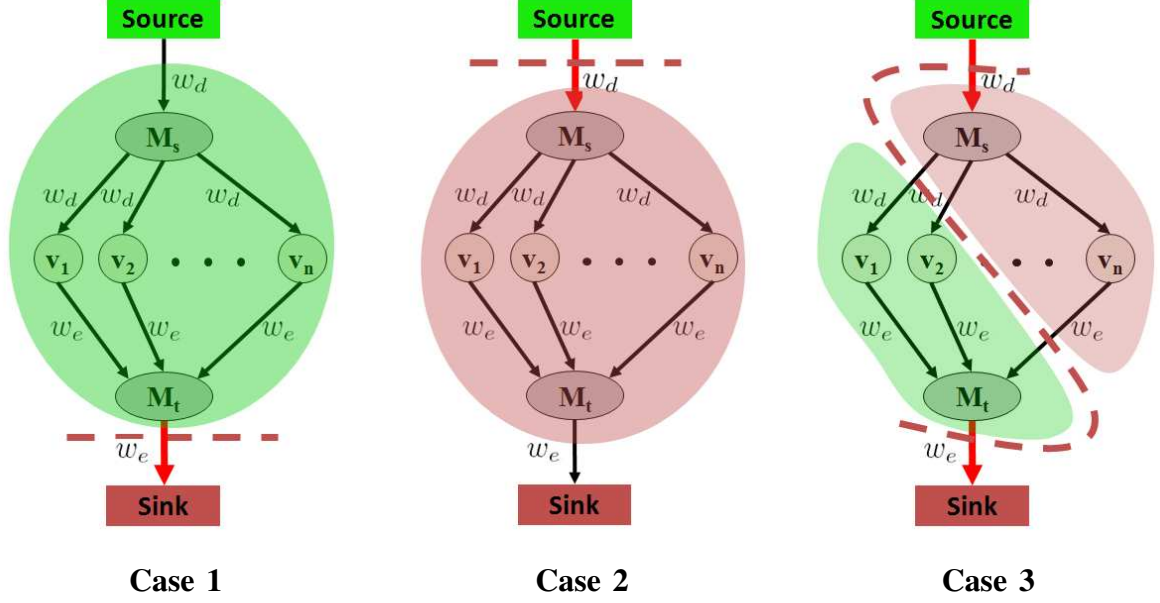


Fig. 2. The figures show which graph edges are cut for different labellings. The edges included in the  $st$ -mincut are shown in bold. Note that only edges going from the source set (shaded green) to the sink set (shaded red) are included in the  $st$ -cut. For instance, in case 3, only two edges are included the  $st$ -cut, the edge between the source and auxiliary node  $M_s$ , and the one between the sink and the auxiliary node  $M_t$ .

Without loss of generality, we assume  $\mathbf{t}_c = \{t_1, t_2, \dots, t_n\}$ . Fig. 1 shows the graph construction corresponding to the above values of the clique potential. Here, the node  $v_i$  corresponds to move variable  $t_i$ . In other words, after the computation of the  $st$ -mincut if  $v_i$  is connected to the source (i.e. it belongs to the source set) then  $t_i = 0$  and if  $v_i$  is connected to the sink (i.e. it belongs to the sink set) then  $t_i = 1$ . 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 (38)).

In order to show that graph constructed above corresponds to the clique potential in equation (37) (plus the constant  $\kappa$ ) we consider three cases:

- 1)  $t_i = 0, \forall i \in c$  : 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. See case 1, figure 2.
- 2)  $t_i = 1, \forall i \in c$  : In this case, the  $st$ -mincut corresponds to the edge connecting the source

with  $M_s$  which has a cost  $w_d = \gamma_\beta + \kappa$ . See case 2, figure 2.

- 3) 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 (39)). See case 3, figure 2.

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. The different cases are illustrated in figure 2.

**$\alpha$ -expansion:** Given a clique  $\mathbf{x}_c$ , our aim is to find the optimal  $\alpha$ -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  $\alpha$ -expansion move  $\mathbf{t}_c$  is given by

$$\psi_c(T_\alpha(\mathbf{x}_c, \mathbf{t}_c)) = \begin{cases} \gamma & \text{if } t_i = 0, \forall i \in c, \\ \gamma_\alpha & \text{if } t_i = 1, \forall i \in c, \\ \gamma_{\max} & \text{otherwise,} \end{cases} \quad (40)$$

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 (37). Therefore, it can be represented using a graph by adding a constant  $\kappa = \gamma_{\max} - \gamma_\alpha - \gamma$ . This proves that the optimal  $\alpha$ -expansion move can be obtained using an st-mincut operation.

## V. 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<sup>2</sup>. This problem can be stated as follows. Given a set of distinct textures (e.g. a dictionary of RGB patches or histograms of textons [26]) 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. 4).

The above problem can be formulated within a probabilistic framework using a CRF [15]. A CRF represents the conditional distribution of a set of random variables  $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$

<sup>2</sup>Our forthcoming work also demonstrates the effectiveness of  $\mathcal{P}^n$  functions on other applications.

given the data  $\mathbf{D}$ . These variables are defined over a lattice  $\mathcal{V} = \{1, 2, \dots, N\}$ . Each of the variables can take one label  $x_i \in \{1, 2, \dots, N_s\}$ . In our case,  $N_s$  is the number of distinct object classes, a variable  $X_i$  represents a pixel  $\mathbf{D}_i$  and  $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$  describes a segmentation. The MAP segmentation solution can be obtained by minimizing the corresponding Gibbs energy.

### A. 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], [4], [23]. In this case, the Gibbs energy of the CRF is of the form:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j), \quad (41)$$

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  $\mathcal{H}_a, a = 1, \dots, N_s$  of the segments as

$$\psi_i(x_i) = -\log p(\mathbf{D}_i | \mathcal{H}_a), \text{ when } x_i = a. \quad (42)$$

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.

$$\psi_{ij}(x_i, x_j) = \begin{cases} \lambda_1 + \lambda_2 \exp\left(\frac{-g^2(i,j)}{2\sigma^2}\right) & \text{if } x_i \neq x_j, \\ 0 & \text{if } x_i = x_j, \end{cases} \quad (43)$$

where  $\lambda_1$ ,  $\lambda_2$  and  $\sigma$  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 [4] for details. Note that the pairwise potentials  $\psi_{ij}(x_i, x_j)$  form a metric. Hence, the energy function in equation (41) can be minimized using both  $\alpha\beta$ -swap and  $\alpha$ -expansion algorithms.

### B. 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], [28]. In this work, we represent the texture of each object class  $s \in \{1, 2, \dots, N_s\}$  using a dictionary  $\mathbf{P}_s$  of  $N_p \times N_p$  RGB patches. Note, however, that our framework is independent of

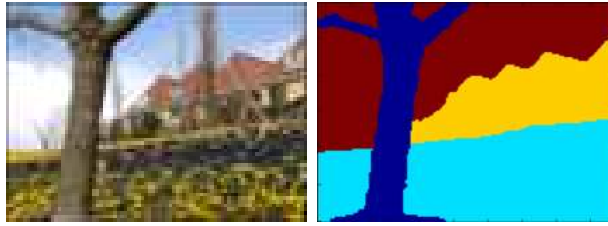


Fig. 3. Segmented keyframe of the garden sequence. The left image shows the keyframe while the right image shows the corresponding segmentation provided by the user. The four different colours indicate pixels belonging to the four segments namely sky, house, garden and tree.

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

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j) + \sum_{c \in \mathcal{C}} \psi_c(\mathbf{x}_c), \quad (44)$$

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  $|\mathcal{C}| = O(n)$ . The unary potentials  $\psi_i(x_i)$  and the pairwise potentials  $\psi_{ij}(x_i, x_j)$  are given by equations (42) and (43) 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.

$$\psi_c(\mathbf{x}_c) = \begin{cases} \lambda_3 G(c, s) & \text{if } x_i = s, \forall i \in c, \\ \lambda_4 & \text{otherwise.} \end{cases} \quad (45)$$

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 is 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  $\alpha$ -expansion algorithms as described in section IV-C.

### C. Segmentation 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

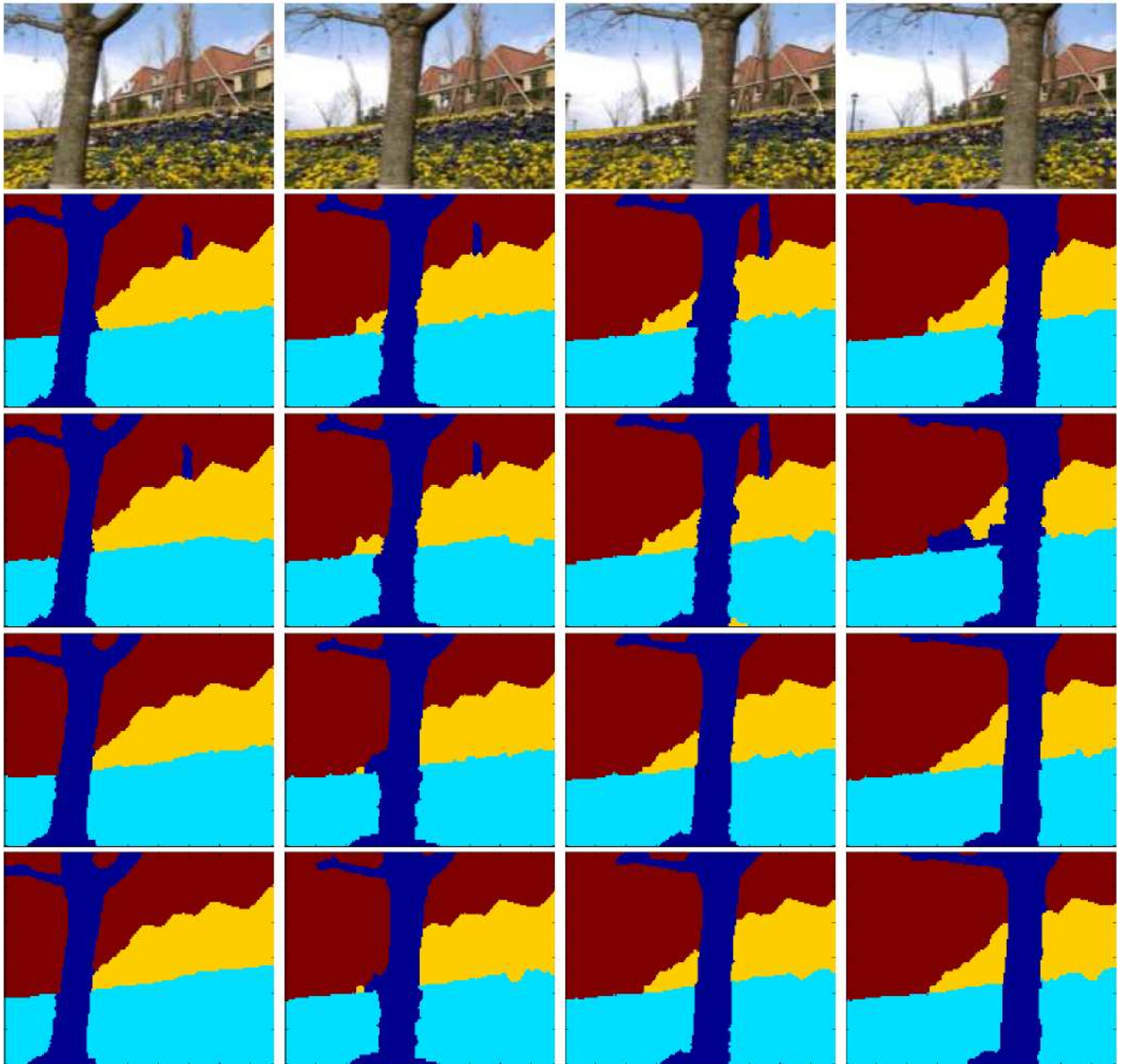


Fig. 4. *Qualitative texture segmentation results of the garden sequence. The first row shows four frames of the garden sequence. The second row shows the segmentation obtained by minimizing the energy of the **pairwise** CRF (in equation (41)) using the  $\alpha\beta$ -swap algorithm. The four different colours indicate the four segments. The segmentations obtained using  $\alpha$ -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 (44)) using the  $\alpha\beta$ -swap algorithm. The fifth row shows the results obtained using the  $\alpha$ -expansion algorithm to minimize the energy in equation (44). The use of higher order cliques results in improved segmentations which are quite close to the ground truth.*



Fig. 5. The figure shows a keyframe of the ‘Dayton’ sequence, and the corresponding user specified segmentation. This segmentation was used to learn the patch dictionaries for the three segments namely background, human1 and human2.

$\mathbf{P}_s$ . The  $\alpha\beta$ -swap and  $\alpha$ -expansion algorithms were used to perform segmentation on the other frames. In all our experiments, we used patches of size  $4 \times 4$ , 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. 3 shows the segmented keyframe of the well-known garden sequence. Fig. 4 (row 2) shows the segmentation obtained for four frames by minimizing the energy function of the pairwise CRF (defined in equation (41)) using the  $\alpha\beta$ -swap algorithm. Note that these frames are different from the keyframe (see Fig. 4 (row 1)). The results obtained by the  $\alpha$ -expansion algorithm are shown in Fig. 4 (row 3). The  $\alpha$ -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. 4 (row 4) shows the segmentations obtained when the energy function of the higher order CRF (defined in equation (44)) is minimized using  $\alpha\beta$ -swap. Fig. 4 (row 5) shows the results obtained using the  $\alpha$ -expansion algorithm. On average,  $\alpha$ -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  $\alpha$ -expansion, the use of higher order cliques provides more accurate segmentation than the pairwise CRF formulation.

Fig. 5 shows another example of a segmented keyframe from a video sequence. The segmentations obtained for four frames of this video are shown in Fig. 6. 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.

**Interactive Image Segmentation:** The higher order CRF model defined above can also be used for interactive image segmentation [4], [23]. In this problem scenario, the user marks certain

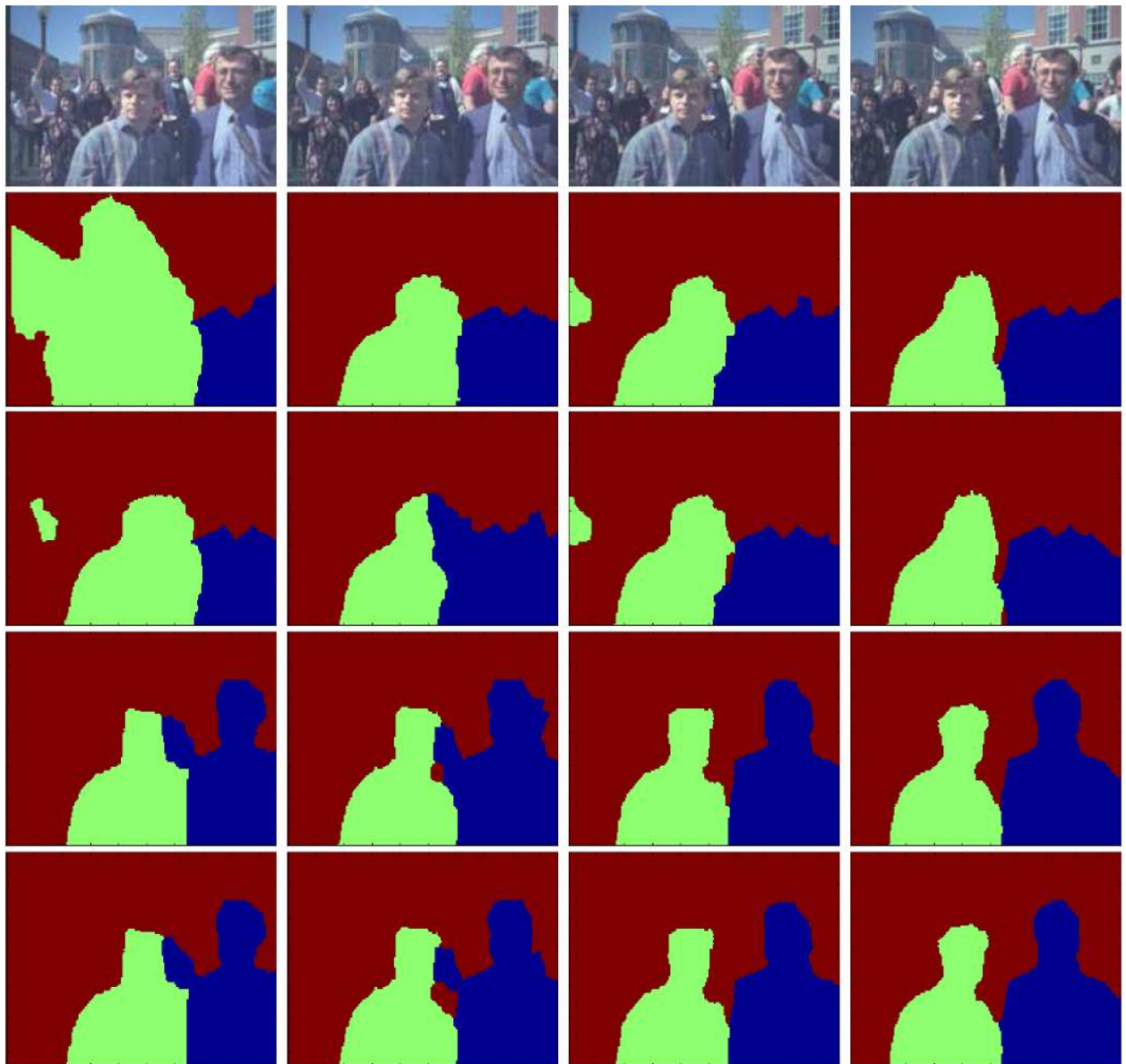


Fig. 6. Segmentation results of the 'Dayton' sequence. Rows 2 and 3 show the results obtained for the frames shown in row 1 by minimizing the energy function in equation (41) using  $\alpha\beta$ -swap and  $\alpha$ -expansion respectively. Row 4 and 5 show the segmentations obtained by minimizing the energy in equation (44) using  $\alpha\beta$ -swap and  $\alpha$ -expansion respectively. The use of higher order cliques defined on patches results in improved segmentations of the two humans.

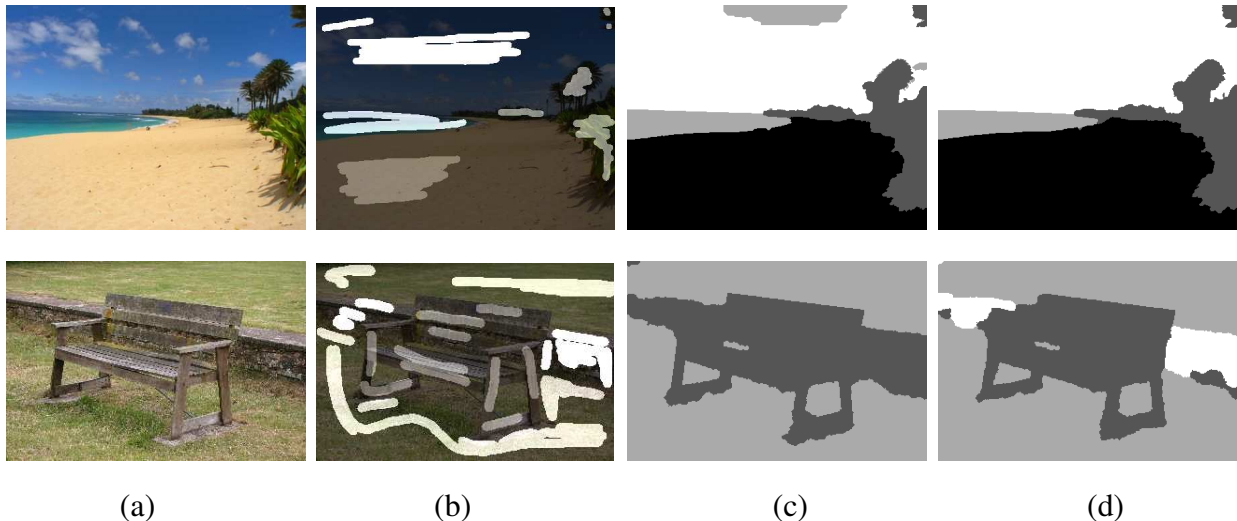


Fig. 7. *Interactive image segmentation with texture based higher order potentials. (a) Two images from the Microsoft Research Cambridge object segmentation dataset. (b) The user made brush strokes which were used to learn colour models and patch dictionaries for the different segments. (c) and (d) Segmentation results obtained by using the alpha expansion algorithm for performing inference in the pairwise CRF and the patch based the higher order CRF respectively. It can be observed that the patch based higher order random field model is able to disambiguate textures having similar colour (sky/water, bench/wall) and thus produces a more accurate segmentation result.*

regions in the image to belong to a particular segment. The pixels constituting these regions are used to learn colour models and patch dictionaries for the respective segments. These models are then used to compute the potentials of the higher order conditional random field model as described earlier. The final segmentation result is obtained by performing inference in this model using the alpha expansion algorithm. Some results of our experiments are shown in figure 7. It can be observed that the use of patch based higher order potentials results in better segmentations.

## VI. PLANARITY PRESERVING CLIQUE POTENTIALS

In this paper we characterized a class of higher order potentials for which the optimal expansion and swap moves can be computed in polynomial time. There are many potential functions for whom it is hard to compute the optimal expansion and swap moves. Potentials which encourage planarity in disparity estimation belong to this class.

Most approaches for disparity estimation assume a Potts model prior on the disparities of neighbouring pixels. This prior favours regions of constant disparity and penalizes discontinuities

in the disparity map. This has the severe side-effect of assigning a high cost to planar regions in the scene which are not orthogonal to the camera-axis.

The above problem can be rectified by using a higher order clique potential which is *planarity preserving*. We define this potential as follows. Consider a higher order clique consisting of three variables  $X_1, X_2$ , and  $X_3$  which can take a disparity label from the set  $\mathcal{L} = \{1, 2, \dots, k\}$ . As before we will use  $x_i$  to denote a labelling of variable  $X_i$ . We say that the clique potential function  $\psi_c()$  is planarity preserving if:

$$\psi_c(x_1, x_2, x_3) = \begin{cases} \gamma_1 & \text{if } |x_2 - x_1| = |x_3 - x_2| \\ \gamma_2 & \text{otherwise.} \end{cases} \quad (46)$$

where  $\gamma_1 < \gamma_2$ .

We will now show that the expansion move energy for this family of clique potentials is not always submodular, and hence cannot always be exactly minimized using algorithms for submodular function minimization.

*Statement 1:* The expansion move energy for the clique potential  $\psi_c$  of the form (46) is non-submodular for some  $\alpha$  and configurations  $\mathbf{x}$ .

Consider the configuration  $\{x_1, x_2, x_3\} = \{1, 2, 1\}$  on which we want to perform a  $\alpha$ -expansion move with  $\alpha = 3$ . The move energy  $E_m$  is a function of the three move variables  $t_1, t_2$ , and  $t_3$  and is defined as:

$$E_m(t_1, t_2, t_3) = E_m^{123} = \begin{array}{|c|c|} \hline E_m(0, 0, 0) & E_m(0, 0, 1) \\ \hline E_m(0, 1, 0) & E_m(0, 1, 1) \\ \hline E_m(1, 0, 0) & E_m(1, 0, 1) \\ \hline E_m(1, 1, 1) & E_m(1, 1, 1) \\ \hline \end{array}$$

which can be written in terms of  $\psi_c$  as:

$$E_m^{123} = \begin{array}{|c|c|} \hline \psi_c(T_\alpha(\{0, 0, 0\}, \mathbf{x})) & \psi_c(T_\alpha(\{0, 0, 1\}, \mathbf{x})) \\ \hline \psi_c(T_\alpha(\{0, 1, 0\}, \mathbf{x})) & \psi_c(T_\alpha(\{0, 1, 1\}, \mathbf{x})) \\ \hline \psi_c(T_\alpha(\{1, 0, 0\}, \mathbf{x})) & \psi_c(T_\alpha(\{1, 0, 1\}, \mathbf{x})) \\ \hline \psi_c(T_\alpha(\{1, 1, 0\}, \mathbf{x})) & \psi_c(T_\alpha(\{1, 1, 1\}, \mathbf{x})) \\ \hline \end{array}$$

For  $\alpha = 3$  and  $\mathbf{x} = \{1, 2, 1\}$  this becomes:

$\psi_c(1, 2, 1)$	$\psi_c(1, 2, 3)$	=	$\gamma_2$	$\gamma_1$
$\psi_c(1, 3, 1)$	$\psi_c(1, 3, 3)$		$\gamma_2$	$\gamma_2$
$\psi_c(3, 2, 1)$	$\psi_c(3, 2, 3)$		$\gamma_1$	$\gamma_2$
$\psi_c(3, 3, 1)$	$\psi_c(3, 3, 3)$		$\gamma_2$	$\gamma_1$

From the definition of submodularity all projections of the move energy  $E_m^{123}$  need to be submodular. Consider the submodularity constraints of the projection  $E_m(0, t_2, t_3)$ :

$$E_m(0, 0, 0) + E_m(0, 1, 1) \leq E_m(0, 0, 1) + E_m(0, 1, 0) \quad (47)$$

or  $\gamma_2 + \gamma_2 \leq \gamma_1 + \gamma_2$ . This constraint is not satisfied as  $\gamma_1 < \gamma_2$ . A similar result can also be proved for  $\alpha\beta$ -swap moves considering the configuration  $\mathbf{x} = \{1, 2, 1\}$  and the 1, 3-swap move.

We have shown that planarity preserving potentials result in non-submodular move energies which cannot be exactly minimized using graph cuts. These move energy functions can be approximately minimized using methods such as [3], [24] which work by solving the roof-dual relaxation of the integer programming problem. This approach has been recently used to obtain promising results for problems such as novel view synthesis [30].

## VII. CONCLUSIONS

In the previous sections, 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 as has been shown by the recent work on object recognition and segmentation [10].

We conclude with the observation that the optimal moves for many interesting clique potentials are hard to compute. Moves for such functions can be computed using recently proposed heuristics for minimizing non-submodular functions [3], [24]. Although these methods are not guaranteed to always produce the optimal move, they have been shown to produce promising results for problems such as novel view synthesis [30].

## VIII. ACKNOWLEDGEMENTS

We thank the associate editor Prof. Ramin Zabih and the anonymous reviewers for numerous suggestions which we believe improved the presentation of the paper. 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.

### APPENDIX I

#### PROOF OF LEMMA

**Lemma 1:** For a continuous function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$y_1 + y_2 \geq y_3 + y_4 \implies f(y_1) + f(y_2) \geq f(y_3) + f(y_4) \quad (48)$$

for all  $y_1, y_2, y_3, y_4 \in \mathbb{R}$  if, and only if,  $f(\cdot)$  is linear.

*Proof:* We only prove the ‘only if’ part. The proof for the forward implication (‘if’) is trivial.

$$x + \epsilon \geq (x - \epsilon) + 2\epsilon \quad (49)$$

$$f(x) + f(\epsilon) \geq f(x - \epsilon) + f(2\epsilon) \quad (50)$$

$$f(x) - f(x - \epsilon) \geq f(2\epsilon) - f(\epsilon) \quad (51)$$

Similarly, starting from  $x + \epsilon \leq (x - \epsilon) + 2\epsilon$ , we get

$$f(x) - f(x - \epsilon) \leq f(2\epsilon) - f(\epsilon). \quad (52)$$

From equations (51) and (52) we get:

$$f(x) - f(x - \epsilon) = f(2\epsilon) - f(\epsilon). \quad (53)$$

Taking limits with  $\epsilon \rightarrow 0$  we get the condition that the derivative (slope) of the function is constant. Hence,  $f(\cdot)$  is linear. ■

### 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.
- [2] E. Boros and P. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155–225, 2002.
- [3] E. Boros, P. L. Hammer, and G. Tavares. Local search heuristics for quadratic unconstrained binary optimization (qubo). *J. Heuristics*, 13(2):99–132, 2007.

- [4] 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.
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.
- [6] D. Freedman and P. Drineas. Energy minimization via graph cuts: Settling what is possible. In *CVPR*, 2005.
- [7] H. Ishikawa. Exact optimization for markov random fields with convex priors. *PAMI*, 25:1333–1336, October 2003.
- [8] 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.
- [9] P. Kohli, M. Kumar, and P. Torr. P3 & beyond: Solving energies with higher order cliques. In *CVPR*, 2007.
- [10] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. In *CVPR*, 2008.
- [11] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10), 2006.
- [12] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts?. *PAMI*, 26(2):147–159, 2004.
- [13] N. Komodakis and G. Tziritas. A new framework for approximate labeling via graph cuts. In *ICCV*, 2005.
- [14] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Extending pictorial structures for object recognition. In *BMVC*, pages II: 789–798, 2004.
- [15] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labelling sequence data. In *ICML*, 2001.
- [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.
- [17] T. Leung and J. Malik. Recognizing surfaces using three-dimensional textons. In *ICCV*, pages 1010–1017, 1999.
- [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.
- [19] A. Neubeck, A. Zalesny, and L. van Gool. 3d texture reconstruction from extensive BTF data. In *Texture*, pages 13–18, 2005.
- [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.
- [21] S. Ramalingam, P. Kohli, A. Karteek, and P. Torr. Exact inference in multi-label crfs with higher order cliques. In *CVPR*, 2008.
- [22] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *CVPR*, pages 860–867, 2005.
- [23] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: interactive foreground extraction using iterated graph cuts. In *SIGGRAPH*, pages 309–314, 2004.
- [24] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary mrfs via extended roof duality. In *CVPR*, 2007.
- [25] D. Schlesinger and B. Flach. Transforming an arbitrary minsum problem into a binary one. Technical Report TUD-F106-01, Dresden University of Technology, 2006.
- [26] F. Schroff, A. Criminisi, and A. Zisserman. Single-histogram class models for image segmentation. In *ICVGIP*, 2006.
- [27] 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 with smoothness-based priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(6):1068–1080, 2008.
- [28] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *CVPR*, pages 691–698, 2003.
- [29] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Tree-based reparameterization for approximate inference on loopy graphs. In *NIPS*, pages 1001–1008, 2001.

- [30] O. Woodford, A. Fitzgibbon, P. Torr, and I. Reid. Global stereo reconstruction under second order smoothness priors. In *CVPR*, 2008.
- [31] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propagation. In *NIPS*, pages 689–695, 2000.