Alphabet SOUP: A Framework for Approximate Energy Minimization

Stephen Gould
Dept. of Electrical Engineering
Stanford University
sgould@stanford.edu

Fernando Amat
Dept. of Electrical Engineering
Stanford University
famat@stanford.edu

Daphne Koller
Dept. of Computer Science
Stanford University
koller@cs.stanford.edu

Abstract

Many problems in computer vision can be modeled using conditional Markov random fields (CRF). Since finding the maximum a posteriori (MAP) solution in such models is NP-hard, much attention in recent years has been placed on finding good approximate solutions. In particular, graph-cut based algorithms, such as α-expansion, are tremendously successful at solving problems with regular potentials. However, for arbitrary energy functions, message passing algorithms, such as max-product belief propagation, are still the only resort.

In this paper we describe a general framework for finding approximate MAP solutions of arbitrary energy functions. Our algorithm (called Alphabet SOUP *for Sequential Optimization for Unrestricted Potentials) performs a search over variable assignments by iteratively solving subproblems over a reduced state-space. We provide a theoretical guarantee on the quality of the solution when the inner loop of our algorithm is solved exactly. We show that this approach greatly improves the efficiency of inference and achieves lower energy solutions for a broad range of vision problems.*

1. Introduction

Many problems in computer vision can be modeled using conditional Markov random fields (CRFs). Solving these problems amounts to maximum a posteriori (MAP) inference, or finding an assignment to each variable that jointly minimizes the energy function (maximizes the probability) defined by the model. Although MAP inference for a general CRF is NP-hard, efficient algorithms exist for some special cases. One important case is that of pairwise binary CRFs with regular potentials, a class that can be solved efficiently using graph-cut-based algorithms. Inspired by this, a number of works have attempted to develop efficient approximation algorithms for the non-binary case. Notably, the α-expansion search method of Veksler *et al*. [26, 1] can be applied to problems with pairwise regular[1] energy functions and has been shown in empirical studies [25] to produce solutions that are near optimal. Thus, for the special case of regular energies, the problem of MAP inference is essentially solved.

Regular energies, and the associated minimization algorithms, are used ubiquitously in addressing early vision tasks, such as dense stereo, image denoising, binary image segmentation, etc. [25], where one often uses a simple (pairwise) smoothness prior between neighboring pixels in a 2D grid. However, as noted Szeliski *et al*. [25], the energy for the groundtruth assignment is often worse than the energy-optimizing assignment, indicating that these simple energy functions fail to model important aspects of the problem. The distance between tractable and useful models becomes even more severe when we use CRFs to model mid-level and high-level vision tasks, such as multi-class image segmentation [9, 20], joint segmentation and detection [15] and 3D reasoning from monocular images [8]. These tasks, while usually having fewer variables than their early vision cousins, have significantly more difficult energy functions, which often include high-order terms, non-grid neighborhoods and heterogeneous variables.

Thus, for many vision applications, the CRFs that adequately capture the important properties of the problem are unlikely to be regular, and therefore are not amenable to the use of the highly-efficient graph-cut-based algorithms. Currently, the only general purpose methods for solving problems with arbitrary energy functions are message passing algorithms such as max-product (MP) belief propagation [19], or its convex variants, such as tree-reweighted message-passing (TRW) [28, 10] or GEMPLP [5]. Unfortunately, these algorithms are often very slow to converge, and cannot handle graphs with very large value spaces. Indeed, as noted in [9], "the lack of efficient algorithms for performing inference in these [higher-order] models has limited their applicability."

In this paper, we aim to meet this challenge, by providing a flexible framework that can produce good approximate solutions and that can scale to accommodate available computing resources and problem complexity. Briefly, we propose a method, called *Alphabet SOUP* (Sequential Op-

[1] Here the regularity condition is on the energy function defined by the α-expansion moves, i.e., $\theta_{ij}(\alpha,\alpha)+\theta_{ij}(\beta,\gamma) \leq \theta_{ij}(\beta,\alpha)+\theta_{ij}(\alpha,\gamma)$.

timization for Unrestricted Potentials), that performs an iterative search over variable assignments. The method performs large global moves in the space by (temporarily) reducing the state-space for each variable, and finding the minimum energy assignment over the reduced state-space. The method is agnostic to the algorithm used for this optimization step, allowing the algorithm best-suited to the particular energy function to be used.

Our method can be viewed as a generalization of the α-expansion search method of Veksler *et al*. [26, 1], which also iteratively proposes steps based on optimizing the energy over a reduced state-space for each variable. However, our more general method is also applicable to CRFs with higher-order cliques and arbitrary energy functions. Our method also allows us to consider a much larger subspace during each iteration of the search, enabling the algorithm to make larger global moves.

Our contributions are threefold: First, we propose a wrapper method for performing approximate MAP inference in graphical models which can be scaled to accommodate different problem sizes and processing limitations. Second, we provide optimality guarantees when the inner loop of our method is exact. Last, we show how the subsets required by our method can be chosen and validate our approach on various contemporary problems. In many cases, our method results in lower energies than were achieved by the methods reported in the literature.

Finally, we note that our Alphabet SOUP method is a general purpose energy minimization technique and not restricted to vision problems. For example, CRFs were first introduced in modeling natural language [16] where they provide state-of-the-art solutions for problems ranging from named-entity recognition to information extraction. They have also been used with great success in computational biology, in applications that include 3D protein-structure prediction [30] and inferring the architecture of cellular networks.

2. Background and Related Work

A Markov random field (MRF) defines a probability distribution $\mathbf{P}(\mathcal{X}) = \frac{1}{Z}\exp\{-\sum_c \theta_c(\boldsymbol{X}_c)\}$ over discrete random variables $\mathcal{X} = \{X_1, \ldots, X_n\}$, where each variable can take on values in some *domain* $\mathbf{dom}(X_i)$. The distribution is parameterized by real-valued potential functions $\theta_c(\boldsymbol{X}_c)$ over sets of variables, or cliques, $\boldsymbol{X}_c \subseteq \mathcal{X}$. The potentials represent a relative preference for every assignment to the variables in the clique $\boldsymbol{X}_c$. For example, in the standard Potts model, a pairwise potential $\theta_{ij}(X_i, X_j)$ assigns a uniform penalty for $X_i \neq X_j$ and no penalty otherwise. The term $E(\boldsymbol{x}) = \sum_c \theta_c(\boldsymbol{x}_c)$ is called the *energy* and the MAP assignment for $\mathbf{P}(\mathcal{X})$ can be found by solving the problem:

$$\begin{array}{ll} \text{minimize} & E(\boldsymbol{x}) = \sum_c \theta_c(\boldsymbol{x}_c) \\ \text{subject to} & x_i \in \mathbf{dom}(X_i) \qquad \forall X_i \in \mathcal{X} \end{array} \tag{1}$$

A large body of literature exists covering MAP inference; here, we provide only a very brief review. We note that Szeliski *et al*. [25] provides a review of different energy minimization methods for computer vision, and a quantitative comparison on a number of benchmark vision tasks.

One of the earliest energy-minimization methods is the still-popular max-product (MP) belief propagation [19]. Here, messages are sent between nodes in the MRF indicating a node's preference for the assignment of its neighbor. Each node accumulates messages from all of its neighbors and maintains a belief (distribution) over possible assignments. The algorithm iterates until beliefs stop changing (or until a maximum number of messages have been sent). The joint MAP assignment is discovered by taking the assignment which locally maximizes each belief.

A different approach is based on viewing the MAP inference problem of Eq. 1 as an integer programming optimization problem, and solving its linear programming (LP) relaxation. Although solving the linear program directly is generally infeasible, several approaches use message-passing-like algorithms to solve its dual; some of these methods are not guaranteed to converge to the dual-optimal solution [28, 10, 5] whereas more recent methods [24, 13] do provide such guarantees. An important advantage of these methods is that, due to the properties of linear programming duality, they provide a lower bound on the energy function. This lower bound can be used to guide the addition of consistency constraints and result in an optimal solution [24]. However, these methods have limited applicability, as they are only usable when the entire problem can be fit in main memory, and are therefore inapplicable to problems where the domain size of the variables is large, or where cliques involve a large number of variables.

In the context of computer vision problems, significant attention has been given to graph-cut based algorithms [6, 26, 1, 12, 4, 25] which have been shown to perform exceptionally well on large grid-structured problems with (regular pairwise) smoothness priors, i.e., problems of the form: $\mathbf{P}(\mathcal{X}) = \frac{1}{Z}\exp\left\{-\sum_{(i,j)} \theta_{ij}(X_i, X_j)\right\}$ where a term is included for every pair of adjacent variables (i, j) and $\theta_{ij}(X_i, X_j)$ is assumed to be a *metric*, encoding a preference for adjacent variables to take on similar values. When the problem is over binary-valued variables with so-called regular potentials, these methods obtain the global optimum. For non-binary problems, i.e., where each variable can be assigned a value from a larger label space $\mathcal{L}$, a search algorithm is generally used, with graph-cut methods providing the optimal move in some constrained search space. One such method, which is closely related to ours, is the α-expansion algorithm [26, 1]. The algorithm maintains a current best joint assignment and iterates over labels $\alpha \in \mathcal{L}$ trying to find a better assignment by allowing variables to either keep their current assignment or change to α. This is called an α-*expansion* move. The algorithm cycles

until no further improvement to the objective can be made. The solution is a local minimum in the sense that no single α-expansion move can result in a lower energy. Here, a global optimum is not guaranteed, but the approach seems to work very well in practice.

There are two main problems with the basic α-expansion algorithm described above. First, it can only be used on pairwise MRFs with regular potentials (and hence also limited to MRFs with homogeneous variables). Second, when the cardinality of each variable is large, it needs many iterations and risks getting stuck in local minima.

Several works attempt to address the first of these issues by extending the graph-cut approach to non-regular potentials. One approach is to approximately solve the pairwise binary MRF required by α-expansion using algorithms such as quadratic pseudo-binary optimization (QPBO) [11, 21]; this method is only applicable when most of the potentials are regular. Another approach is to develop algorithms for special-case energy functions, such as truncated convex priors [27]. While all of these methods are effective for some problems, they are limited to pairwise potentials and do not address the problem of optimizing general energy functions.

Other works aim to address the issue of large value spaces by reducing the set of labels considered. The novel fusion-move approach [18] makes moves by combining two proposed solutions $\boldsymbol{x}^0$ and $\boldsymbol{x}^1$. These fusion-moves usually result in non-regular energies and so the algorithm resorts to approximate inference (e.g., QPBO). Our work generalizes this approach by allowing the *fusing* of multiple proposed solutions in a single search step. In addition, we provide theoretical guarantees when the search steps are exact.

Recent work in computer vision has started to make use of higher-order cliques and problems over heterogeneous variables. Lan *et al*. [17] showed that for the problem of image denoising the value space for each variable can be pruned reliably (by examining its local neighborhood) making belief propagation tractable. However, their method is not general and does not provide any guarantee on the quality of the solution. Other recent work [20] shows how to transform multi-label high-order energy functions into second-order binary ones which can then be solved by approximation techniques. It is not clear the extent to which the approximation at the binary level affects the multi-label result. Furthermore, for problems with large variable domains, the transformation to binary can be prohibitive.

3. Alphabet SOUP

We now describe a new algorithm for approximate MAP inference. Like α-expansion, we aim to optimize the assignment by performing a search over the value space: we maintain a current best joint assignment to the random variables, which we modify by searching over a space of possible moves. However we will consider a much richer set of moves than α-expansion.

3.1. γ-expansion Moves

Let $A_i^\gamma \subseteq \mathbf{dom}\,(X_i)$ be a subset of the domain for each variable X_i. We define a γ-*expansion* move to be a mapping for each variable X_i from its current value x_i to a value $\hat{x}_i \in A_i^\gamma \cup \{x_i\}$. Our goal is to find the assignment that is a γ-expansion move that has minimum energy. We do this by defining a new energy function $E^\gamma(\hat{\boldsymbol{x}}; \boldsymbol{x}) = \sum_c \theta_c^\gamma(\hat{\boldsymbol{x}}_c)$ over the restricted domain for each variable $A_i^\gamma \cup \{x_i\}$ where we construct each potential $\theta_c^\gamma(\hat{\boldsymbol{x}}_c)$ as

$$\theta_c^\gamma(\hat{\boldsymbol{x}}_c) = \begin{cases} \theta_c(\hat{\boldsymbol{x}}_c) & \text{if } \forall X_i \in \boldsymbol{X}_c : \hat{x}_i \in A_i^\gamma \cup \{x_i\} \\ \infty & \text{otherwise} \end{cases} \quad (2)$$

That is, for every value in the potential, we either copy the value if it corresponds to an assignment in our restricted state-space, or set it to infinity otherwise. This reduction operation is illustrated in Figure 1. Notice that the entries corresponding to assignments outside of the restricted state-space can be removed from the potential, creating a potential whose size is a factor of $\prod_{X_i \in \boldsymbol{X}_c} \frac{|\mathbf{dom}(X_i)|}{|A_i^\gamma \cup \{x_i\}|}$ smaller. This decrease in size can lead to substantial speed improvements by allowing potentials to fit in memory.

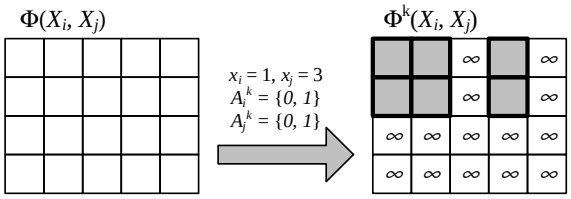

Figure 1: Illustration of restricting a potential for evaluating the optimal γ-expansion move. In this example the current best assignment to (X_i, X_j) is $(1, 3)$. If $A_i^\gamma = A_j^\gamma = \{0, 1\}$ then the restricted potential will be over the entries in $\{0, 1\} \times \{0, 1, 3\}$ as shown.

The restricted potential allows us to efficiently find the optimal γ-expansion move:

Observation 3.1. *Finding the MAP assignment for the problem with all potentials restricted to $A_i^\gamma \cup \{x_i\}$ is equivalent to finding the optimal γ-expansion move from $\boldsymbol{x}$.*

3.2. Using γ-Expansion Moves

Based on the notion of a γ-expansion move, we can now define an algorithm that iteratively searches over the space. At each point, we select a particular γ-expansion operation. We then use the energy-minimization algorithm of our choice to find the assignment $\boldsymbol{x}'$ that is the optimal γ-expansion move from $\boldsymbol{x}$ and accept the move if it results in a lower energy, i.e., if $E(\boldsymbol{x}') < E(\boldsymbol{x})$. The algorithm continues until none of the γ-expansion moves that we are willing to consider improves the energy, so that we have found a local optimum in our search space.

To define the algorithm concretely, we need to specify which γ-expansions we want to consider at each iteration.

Most simply, we can statically partition the domain of each variable into K subsets $A_i^1, \ldots, A_i^K$; these subsets need not be disjoint and may even be empty. Then, we define γ_k to be the set $\{A_1^k, \ldots, A_n^k\}$, and iterate over the k's in a round-robin fashion. We note that standard α-expansion is a special case of this static variant of our algorithm, where we select A_i^k to be the singleton set containing the k-th label.

However, our framework also allows substantially greater flexibility, in several dimensions. First, we can select subsets that include more than one value for each variable. For example, it is very common for variable assignments to represent some ordinal values (e.g., disparities in stereo reconstruction). In this case, an obvious partition is to group labels into contiguous ranges. At each iteration, the algorithm chooses to keep the current assignment to each variable X_i or change it to one of the values in the range specified by A_i^k. By allowing overlapping partitions, variables can smoothly move from one ordinal range to the next.

A second dimension of flexibility is our method's ability to choose the subsets to reflect the properties of the energy function. For example, one useful heuristic for choosing expansion moves is to group low energy assignments together, as these are likely to occur in low energy solutions. In particular, when singleton potentials are very strong, we might choose to construct a γ that places in each A_i^γ the assignments to X_i that receive low values in $\theta_i(X_i)$. When pairwise (or higher-order) terms are strong, we can choose to group values that jointly give rise to low energy configurations within individual cliques; that is, if $\theta_{ij}(x_i, x_j)$ is low, then we might put $(x_i, x_j) \in A_i^\gamma \times A_j^\gamma$, for some γ.

A third dimension of flexibility allows us to construct the expansion moves dynamically, based on the current assignment $\boldsymbol{x}$. For example, we might construct our current γ-expansion move so as to include in A_i^γ assignments x_i that are compatible (achieve low energy) with the current assignments to variables $X_j \neq X_i$. This approach is related to the value pruning methods found in the literature [17, 14], but does not require that the values be pruned permanently. In fact, our Alphabet SOUP method can provide a theoretical foundation for these methods: We can use the value pruning techniques to define the γ-expansion moves in early iterations of the algorithm, but then use a covering set of γ's as a final iteration to provide ourselves with the theoretical guarantees as we discuss in Section 3.3 below.

Dynamic construction is also useful when vector-valued variables (e.g., 3D surface normals) are quantized into a discrete label space. Here, a natural search procedure is to apply coordinate descent on each dimension. In this case the partitions are chosen dynamically based on the decoded coordinate value for each assignment. Here, the subsets correspond to all the assignments consistent with the best assignment at hand, while allowing one coordinate to vary.

Many other heuristics are possible, and their development is an interesting direction for further research.

3.3. Theoretical Guarantees

Above, we discussed different options for selecting the set of possible expansion moves that we consider. We now provide a result that shows that, under weak conditions, the local optimality of an assignment $\boldsymbol{x}$ in the γ-expansion space implies a bound on the distance between the energy of $\boldsymbol{x}$ and the optimal energy.

We define a set of γ-expansion moves $\gamma_1, \ldots, \gamma_K$ to be *covering* if, for every $x_i \in \textbf{dom}\,(X_i)$, there exists a γ_k such that $x_i \in A_i^{\gamma_k}$. When each γ-expansion move is optimal, we can make the following guarantee.

Theorem 3.2. *Let $\gamma_1, \ldots, \gamma_K$ be a covering set of moves. Assume that $\theta_c(\boldsymbol{x}_c) \geq 0$ for all cliques c, with equality only if there exists some γ_k such that $x_i \in A_i^{\gamma_k}$ for all variables X_i in the clique. If $\boldsymbol{x}$ is a local optimum relative to $\gamma_1, \ldots, \gamma_K$, then $E(\boldsymbol{x})$ is within a factor of $\lambda\,(\max_c |\boldsymbol{X}_c|)$ of the optimal energy, where*

$$\lambda = \max_{c:|\boldsymbol{X}_c|>1} \left(\frac{\max_{\boldsymbol{x}_c} \theta_c(\boldsymbol{x}_c)}{\min_{\boldsymbol{x}_c:\theta_c(\boldsymbol{x}_c)\neq 0} \theta_c(\boldsymbol{x}_c)} \right) \tag{3}$$

and $|\boldsymbol{X}_c|$ is the number of variables in clique c.

This theorem (see appendix for proof) subsumes the optimality result for α-expansion [26], i.e., that α-expansion returns an assignment that is within a constant factor of the global optimum. For a Potts model, this ensures that the energy is within a factor of two of the optimum.

Many state-of-the-art techniques exist for solving small problems exactly and can be used for the inner loop of our algorithm, e.g., the junction tree algorithm [2] for problems with small treewidth, min-cut [12] for binary problems with regular potentials, or linear programming (LP) relaxation with cluster pursuit [24]. However our algorithm is well defined even if the inner loop is not solved exactly, allowing researchers to use approximate MAP inference algorithms that are appropriate for their problem.

Recall that a major benefit of the class of LP-based methods is that they exploit duality to place a bound on the distance between the energy of the current assignment and the optimal energy. When such a method is used to perform the optimization for the γ-expansion steps in the inner loop of the Alphabet SOUP algorithm, we can provide similar bounds. In particular, Globerson *et al.* [5] show that the dual of the LP relaxation of Eq. 1 can be reformulated as:

$$\begin{array}{ll} \text{maximize} & \sum_s \min_{\boldsymbol{x}_s} \sum_{c \in N(s)} \min_{\boldsymbol{x}_{c \setminus s}} \beta_c^s(\boldsymbol{x}_c) \\ \text{subject to} & \sum_{s \in S(c)} \beta_c^s(\boldsymbol{x}_c) = \theta_c(\boldsymbol{x}_c), \quad \forall c, \boldsymbol{x}_c \end{array} \tag{4}$$

where s enumerates the set of non-empty intersections, or separators, between cliques, and $S(c)$ and $N(s)$ represent the neighborhoods of clique c and separator s, respectively. The $\beta_c^s(\boldsymbol{x}_c)$ are the dual variables to the primal LP constraints. As usual, the dual objective at any feasible assignment provides a lower bound on $E(\boldsymbol{x})$. Globerson *et*

al. show that this dual LP can be solved efficiently using a message-passing algorithm similar to belief propagation.

When using this LP-based approach to solve the restricted optimization defined by a γ-expansion step, the primal is restricted to the value-space defined by each A_i^γ. The effect on the dual is that $\beta_c^s(\boldsymbol{x}_c)$'s corresponding to assignments not allowed by the γ-expansion move can be removed from the objective. The dual optimum of the restricted problem is now not guaranteed to be a feasible point for the original unrestricted dual of Eq. 4. However, we can use the restricted solution to produce a dual-feasible assignment to Eq. 4, which thereby immediately provides a bound on the duality gap for the original problem. Specifically, the solution to the restricted dual provides dual-feasible assignments to the $\beta_c^s(\boldsymbol{x}_c)$ corresponding to $\boldsymbol{x}_c$'s allowed by the γ-expansion move, leaving us only to find a feasible assignment to the remaining $\beta_c^s(\boldsymbol{x}_c)$'s. A simple solution is to split the mass of $\theta_c(\boldsymbol{x}_c)$, giving $\beta_c^s(\boldsymbol{x}_c) = \frac{1}{|S(c)|}\theta_c(\boldsymbol{x}_c)$. The resulting solution is dual feasible and hence provides a bound Δ on the distance between our current assignment's energy and the optimal energy as

$$\delta_c^s(\boldsymbol{x}_s) = \min_{\boldsymbol{x}_{c\setminus s}} \left\{ \left\{ \beta_c^s(\boldsymbol{x}_c) \right\}_{\boldsymbol{x}_c \in \gamma}, \left\{ \tfrac{1}{|S(c)|}\theta_c(\boldsymbol{x}_c) \right\}_{\boldsymbol{x}_c \notin \gamma} \right\} \quad (5)$$

$$\Delta = E(\boldsymbol{x}) - \sum_s \min_{\boldsymbol{x}_s} \sum_{c \in N(s)} \delta_c^s(\boldsymbol{x}_s). \quad (6)$$

An interesting avenue for future research is to investigate using the terms δ_c^s that cause an increase in the dual objective to guide the construction of dynamic expansion moves.

4. Experimental Results

We now provide an Alphabet soup of example problems that can be solved by our framework.

Image Completion and Inpainting. Exemplar-based image completion [3, 14] is a method for filling-in missing parts of an image by copying patches from other parts of the image. Recently, Komodakis and Tzititas [14] formulated the problem as a pairwise MRF over grid locations within the missing region. Briefly, fixed-sized patches (from the observed part of the image) are placed in overlapping fashion on the grid to complete the image. Grid locations around the perimeter of the missing region are assigned a singleton term $\theta_i(X_i)$ measuring the sum-of-square-difference (SSD) between the observed region of the grid location and the candidate patch. Similarly, a pairwise SSD term $\theta_{ij}(X_i, X_j)$ is defined for every neighboring grid location (see [14] for details). This energy function is not metric and hence the α-expansion algorithm cannot be used.

Since patches can be drawn from any location within the observed part of the image, the value space is enormous, e.g., roughly 70,000 for a 320×240 image. Clearly, standard message passing algorithms cannot support a problem of this magnitude. To solve the problem, Komodakis and Tzititas propose a priority-based message scheduling algorithm with label pruning. Their approach is to run a belief propagation algorithm in which messages are scheduled according to the current belief and the value-space for each variable is pruned the first time it sends a message. This greedy approach results in a smaller MRF in which pairwise terms can be computed efficiently. However, the pruned labels are never reconsidered and therefore the approach may result in suboptimal energies.

(i) (ii) $E = 3.37$ (iii) $E = 4.97$ (iv) $E = 2.84$

(a) Bungee jumping example from [3].

(i) (ii) $E = 3.26$ (iii) $E = 3.36$ (iv) $E = 2.03$

(b) Author removed from an image.

Figure 2: Comparison of different methods on image completion task. We use a 7×7 grid spacing; (a) required 280 patches to fill, (b) required 179 patches. Results are annotated with cost (energy) per patch. Our method achieves lower energy than the other more greedy approaches.

Instead of pruning values we applied our method of iterating over subsets of label assignments, but still considering all possible labels in the end. Concretely, we divided the label space (set of possible patches) into non-overlapping sets $\mathcal{L}^k$, each containing 250 patches. During each iteration k, we set $A_i^{\gamma_k}$ for the first variable X_i in each row of the missing grid to $\mathcal{L}^k$. We then set the remaining $A_j^{\gamma_k}$ in each row to take labels from the observed part of the image corresponding to labels in $\mathcal{L}^k$ offset by the same distance as between X_i and X_j. This approach helped provide many low energy pairwise matches.

In our experiments we found the method of Komodakis and Tzititas to be very sensitive to the priority schedule and amount of pruning. This makes it susceptible to the same sort of errors produced by more greedy approaches. We ran a number of trials and report the best energy found.

Our naive implementation did not include any speed ups for computing the SSD terms (e.g., computing them in the frequency domain); thus, running time was dominated by the SSD calculations. Nevertheless, our method only ran about 5 times slower than the competing approach.

Results are shown in Figure 2. Here we compare the en-

ergy per variable for the method of Komodakis and Tzititas, a naive greedy approach, and Alphabet SOUP (using max-product message passing as the inner loop). On both test images, we achieve a lower energy.

Object Detection and Outlining. The task of object outlining involves finding instances of an object class in novel images and providing a precise outline around those objects. We tried out our method on the CRF-based LOOPS model of Heitz *et al*. [7], in which correspondences are found between landmarks on the online of an object (e.g., animal's nose) and image pixels. Their model defines a CRF where the variables are the n landmarks, and their assignments represent options for corresponding image pixels. To allow the use of discrete energy optimization techniques, they consider, for each landmark, the m pixels that give rise to the lowest energy values in the (learned) singleton potentials. As n and m are often reasonably small (around 50 or 60), the LOOPS models fit easily into memory. Nevertheless, due to the dense connectivity of the model, standard belief propagation has trouble performing inference and often fails to converge. The LOOPS method [7] handles this problem by removing weak pairwise relationships to produce a sparse model.

We performed experiments on 42 images of giraffes using a LOOPS model with 60 landmarks (see Figure 3), comparing results from our method with the best singleton landmark locations and those found by the discrete stage of the LOOPS approach. As the singleton potentials are very informative in this application, we partitioned the domain of each variable into subsets according to the score of the singleton potentials, using three candidates in each A_i^γ. We ordered subsets with the lowest energy first, so that a low energy assignment was generally found in the very first iteration, with later iterations serving only to correct a few outliers. In our significantly smaller models, belief propagation had no convergence problems, and so we could use the full LOOPS model with all pairwise interactions, rather than the sparsified version. This approach consistently found solutions that have lower energy and are more visually appealing than the other two.

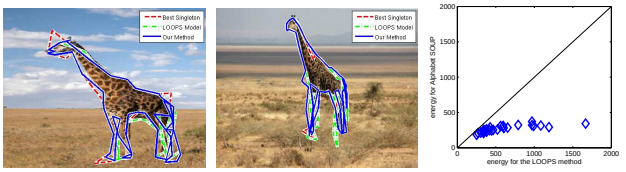

Figure 3: Object outlining using the model of Heitz *et al*. [7] comparing best scoring independent match (red dashed), discrete round of inference in [7] (green dotted), and Alphabet SOUP (blue solid). Our method consistently finds a lower-energy solution than the other two approaches (right).

Surface Reconstruction. One general class of vision applications is that of 3D surface reconstruction from a set of 2D measurements collected from different viewing angles. Here, we can model the surface as a mesh of small patches, where each is parameterized by a 4-dimensional vector defining its (3D) orientation and (1D) radial offset. The singleton potentials generally measure the fit between the patch position and the image(s) obtained from the relevant viewing angles. The pairwise potentials impose a preference for smoothness of the reconstructed surface.

Problem	Energy per Pixel (i)	(ii)	(iii)	Running Time (s) (i)	(ii)	(iii)
(a) 40×43	**0.507**	0.509	0.521	39	1	$< \mathbf{1}$
(b) 112×116	n/a	**0.527**	0.555	∞	23	**5**

Table 1: Results for the 3D surface reconstruction experiments. Shown are results for (i) optimizing all coordinates simultaneously, (ii) coordinate descent over pairs of coordinates, and (iii) coordinate descent over individual coordinates. Problem (a) was over a 40×43 mesh; (b) was over a 112×116 grid (and too large to solve by (i)). Variables were quantized into $(13, 7, 7, 13)$ and $(23, 9, 9, 23)$ bins, respectively.

To allow discrete energy optimization to be applied to this task, we can discretize this 4-dimensional vector. The energy function here is metric and so amenable to the α-expansion algorithm. However, depending on the number of quantization bins per dimension, the state-space for each variable can be very large (often containing around 40,000 values). In these cases, the pairwise smoothness term can be too large to precompute and too expensive to compute on the fly during each iteration of α-expansion. The Alphabet SOUP approach provides a way for reducing the computational cost, by using a coordinate descent variation, with the A_i^γ chosen dynamically, as described in Section 3.2.

We experiment with this approach over the task of reconstructing the outermost surface layer (S-layer) of a bacteria from images obtained by 3D-tomography from cryo-electron microscopes. The S-layer often exhibits geometrical lattice-like 2D structure [23, 22], which provides insight into how the bacteria interacts with its environment. These structures are not easily visible in the raw 2D images, but much more easily discerned in a 3D surface reconstruction.

We evaluated on two different size problems using three approaches: (i) optimizing all coordinates at once, (ii) coordinate descent over pairs of coordinates, and (iii) coordinate descent over individual dimensions. In each case (for our inner loop) we use the α-expansion algorithm using code available online. Results are shown in Table 1.

As expected, optimizing over all coordinates simultaneously obtains the lowest energy. However, very little penalty is paid in terms of energy when performing coordinate descent over pairs of coordinates, while a significant improvement in running time can be gained. In the smaller of our two problems, a 40-fold reduction in running time is obtained, at negligible cost in the energy obtained; in the larger problem, standard α-expansion was simply too large to be solved without coordinate ascent.

Rosetta Protein Design. We also ran our algorithm on some non-vision applications to verify its ability to solve difficult problems consisting of heterogeneous variables

with large domains. We chose the challenging Rosetta Protein Design dataset made available by Yanover *et al*. [29]. The dataset contains 97 models that aim to find the most stable sequence of amino-acids that give rise to a given 3D structure. These problems cannot be solved using standard min-cut approaches since each variable has a different domain. Sontag *et al*. [24] showed that these models could be solved using a state-of-the-art dual message passing algorithm [5] with cluster pursuit. Using their method we were able to find the true MAP solution for 93 of the 97 problems.[2] In Figure 4 we compare the energy and running time of our algorithm with asynchronous max-product belief propagation (MP) and the exact method. We also compare against the dual message passing algorithm, GEMPLP [5], *without* cluster pursuit. Surprisingly, our results show that, not only does Alphabet SOUP run faster (in 91.4% cases) and require less memory than MP or GEMPLP, it also often produces lower energies. In particular, in 96.8% of cases our method is within 5% of the optimal solution.

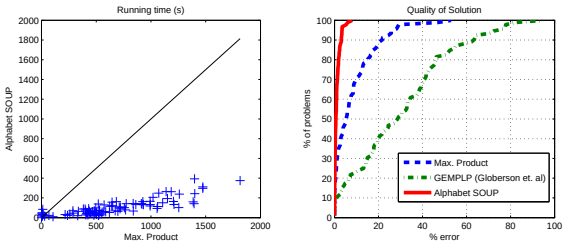

Figure 4: Results for the Rosetta Design Dataset. (left) compares running time in seconds of Alphabet SOUP against async. MP belief propagation. (right) shows the quality of solutions for MP, GEMPLP and Alphabet SOUP compared to the optimum energy. For MP we ran for a maximum of 1000 iterations, most problems failing to converge. For Alphabet SOUP we set the subset size for all variables to 50 and ran asynchronous MP as the inner loop. Mean running time for the exact method [24] was 15 hours.

5. Discussion

In this work, we presented a method for finding approximate solutions to the MAP inference problem for arbitrary energy functions. We can provide an optimality bound on the solution when the inner loop of our method is exact. However, as we showed in our experiments, even when the inner loop cannot be solved exactly, our method still provides several advantages over other approaches. In particular, our method is faster than standard max-product belief propagation, requires significantly less memory, and often produces lower energy solutions.

Perhaps the most interesting directions for further study are in providing more formal foundations for the choice of subsets used for the different variables. First, the empirical observation that problems with smaller domains can be solved more easily deserves more theoretical attention. It would also be interesting to see whether our theoretical bound can be used to provide more formal guidance as to which subsets are likely to give better bounds. Along similar lines, it would be valuable to further explore the connections between these value-based approximations and the linear-program relaxation of the MAP problem, with the goal of providing a bound on the gap between the current solution and the optimal one. Last, the ability to dynamically select expansion moves suggests a *subset pursuit* method, in which subsets are dynamically chosen in a way that facilitates greatest decrease in energy.

Finally, as noted by Szeliski *et al*. [25], groundtruth assignments often fare worse in terms of energy than solutions produced by current state-of-the-art energy minimization techniques, indicating that the energy functions are too simple and fail to model important aspects of the problem. A flexible MAP inference algorithm that caters to large cliques and heterogeneous variables may allow vision researchers to more explore energy functions that are better-suited to their problems.

Acknowledgments. We thank Geremy Heitz for providing the LOOPS data, and Nikos Komodakis and David Sontag for communications on their respective works. This work was supported by the DARPA Transfer Learning program under contract SA4996-10929-4 and the Multidisciplinary University Research Initiative (MURI) under contract number N000140710747.

A. Appendix

Proof of Theorem 3.2. Let $\boldsymbol{x}^\dagger$ be a local minimum in the expansion move space and let $\boldsymbol{x}^\star$ be the global optimum. Fix some k and let $\mathcal{X}_k = \{X_i : x_i^\star \in A_i^{\gamma_k}\}$, i.e., the set of variables whose optimal assignment is within the k-th subset for that variable. We can produce a labeling $\boldsymbol{x}'$ within one γ_k-expansion move from $\boldsymbol{x}^\dagger$ as follows: $x'_i = x_i^\star$ if $X_i \in \mathcal{X}_k$, and $x'_i = x_i^\dagger$ otherwise. Now, since $\boldsymbol{x}^\dagger$ is a local minimum,

$$E(\boldsymbol{x}^\star) \leq E(\boldsymbol{x}^\dagger) \leq E(\boldsymbol{x}') \tag{7}$$

For any set of cliques S, define $E_S(\boldsymbol{x})$ to be the restriction of the energy to that set. Formally, $E_S(\boldsymbol{x}) = \sum_{c \in S} \theta_c(\boldsymbol{x}_c)$.

Define three sets: $I^k = \{c : \boldsymbol{X}_c \subseteq \mathcal{X}_k\}$, the set of all cliques where the variables X_i have their optimal assignment in $A_i^{\gamma_k}$; $B^k = \{c : \boldsymbol{X}_c \cap \mathcal{X}_k \neq \emptyset, \boldsymbol{X}_c \nsubseteq \mathcal{X}_k\}$, the set of all cliques where at least one variable X_i has optimal assignment in $A_i^{\gamma_k}$ and one variable X_j has optimal assignment outside of $A_j^{\gamma_k}$; and $O^k = \{c : \boldsymbol{X}_c \cap \mathcal{X}_k = \emptyset\}$, the set of all cliques where the variables X_i have their optimal assignment outside of $A_i^{\gamma_k}$. For any assignment $\boldsymbol{x}$ we can write $E(\boldsymbol{x}) = E_{I^k}(\boldsymbol{x}) + E_{B^k}(\boldsymbol{x}) + E_{O^k}(\boldsymbol{x})$.

The following is true: $E_{O^k}(\boldsymbol{x}') = E_{O^k}(\boldsymbol{x}^\dagger)$, $E_{I^k}(\boldsymbol{x}') = E_{I^k}(\boldsymbol{x}^\star)$, and $E_{B^k}(\boldsymbol{x}') \leq \lambda E_{B^k}(\boldsymbol{x}^\star)$ where λ is defined in Theorem 3.2. The first two are obvious (and can be seen by summing the relevant $\theta_c(\boldsymbol{x}_c)$). The the last holds because

$$E_{B^k}(\boldsymbol{x}') \leq \max_{\boldsymbol{x}} E_{B^k}(\boldsymbol{x}) \leq \sum_{c \in B^k} \max_{\boldsymbol{x}_c} \theta_c(\boldsymbol{x}_c)$$

[2] The exact method is very computationally intensive and the remaining four problems exceeded the runtime limits (160 hours) on our cluster computer before convergence and so we could not solve them exactly.

$$= \sum_{c \in B^k} \left(\frac{\max_{\boldsymbol{x}_c} \theta_c(\boldsymbol{x}_c)}{\min_{\boldsymbol{x}_c : \theta_c \neq 0} \theta_c(\boldsymbol{x}_c)} \right) \min_{\boldsymbol{x}_c : \theta_c \neq 0} \theta_c(\boldsymbol{x}_c)$$

$$\leq \lambda \sum_{c \in B^k} \min_{\boldsymbol{x}_c : \theta_c \neq 0} \theta_c(\boldsymbol{x}_c) \leq \lambda \sum_{c \in B^k} \theta_c(\boldsymbol{x}_c^\star) = \lambda E_{B^k}(\boldsymbol{x}^\star)$$

where for the last inequality used the fact that for $c \in B^k$ we have $\theta_c(\boldsymbol{x}_c^\star) \neq 0$ by the conditions of our theorem.

Substituting the above into Eq. 7 and applying some simple algebraic manipulation, we have that

$$E_{I^k}(\boldsymbol{x}^\dagger) + E_{B^k}(\boldsymbol{x}^\dagger) \leq E_{I^k}(\boldsymbol{x}^\star) + \lambda E_{B^k}(\boldsymbol{x}^\star).$$

Now consider the case that the $A_i^{\gamma_k}$ are disjoint. Summing over k we have for the left-hand side:

$$\sum_k \Big(\sum_{c \in I^k} \theta_c(\boldsymbol{x}_c^\dagger) + \sum_{c \in B^k} \theta_c(\boldsymbol{x}_c^\dagger) \Big) \geq \sum_{c \in \bigcup_k I^k \cup B^k} \theta_c(\boldsymbol{x}_c^\dagger) = E(\boldsymbol{x}^\dagger)$$

Similarly for the right-hand side:

$$\begin{aligned}
&\left(\textstyle\sum_k \sum_{c \in I^k} \theta_c(\boldsymbol{x}_c^\star)\right) + \lambda \left(\textstyle\sum_k \sum_{c \in B^k} \theta_c(\boldsymbol{x}_c^\star)\right) \\
&\quad \leq \left(\textstyle\sum_k \sum_{c \in I^k} \theta_c(\boldsymbol{x}_c^\star)\right) + \lambda \left(\textstyle\sum_{c \in \bigcup B^k} |\boldsymbol{X}_c| \cdot \theta_c(\boldsymbol{x}_c^\star)\right) \\
&\quad = \left(\textstyle\sum_{c \in \bigcup I^k} \theta_c(\boldsymbol{x}_c^\star) + \sum_{c \in \bigcup B^k} \theta_c(\boldsymbol{x}_c^\star)\right) + \\
&\qquad \left(\textstyle\sum_{c \in \bigcup B^k} (\lambda |\boldsymbol{X}_c| - 1) \cdot \theta_c(\boldsymbol{x}_c^\star)\right) \\
&\quad \leq E(\boldsymbol{x}^\star) + (\lambda \max_c |\boldsymbol{X}_c| - 1) \textstyle\sum_{c \in \bigcup B^k} \theta_c(\boldsymbol{x}_c^\star) \\
&\quad \leq E(\boldsymbol{x}^\star) + (\lambda \max_c |\boldsymbol{X}_c| - 1)\, E(\boldsymbol{x}^\star) \\
&\quad = \lambda \left(\max_c |\boldsymbol{X}_c|\right) E(\boldsymbol{x}^\star)
\end{aligned}$$

where we have used the fact that due to the disjointness of the $A_i^{\gamma_k}$ we cannot have terms appearing in B^k more than $|\boldsymbol{X}_c|$ times. Now for $A_i^{\gamma_k}$ not disjoint, define $\tilde{A}_i^{\gamma_k} = A_i^{\gamma_k} \setminus \bigcup_{l=1}^{k-1} A_i^{\gamma_k}$. The proof above holds for $\tilde{A} i^{\gamma_k}$ which are subsets of the $A_i^{\gamma_k}$ and so holds in general. Thus we have $E(\boldsymbol{x}^\star) \leq E(\boldsymbol{x}^\dagger) \leq \lambda \left(\max_c |\boldsymbol{X}_c|\right) E(\boldsymbol{x}^\star)$. □

References

[1] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In *ICCV*, 1999.

[2] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer-Verlag, 1999.

[3] A. Criminisi, P. Perez, and K. Toyama. Object removal by exemplar-based inpainting. In *CVPR*, 2003.

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

[5] A. Globerson and T. Jaakkola. Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In *NIPS*, 2007.

[6] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society*, 1989.

[7] G. Heitz, G. Elidan, B. Packer, and D. Koller. Shape-based object localization for descriptive classification. In *NIPS*, 2008.

[8] D. Hoiem, A. Stein, A. Efros, and M. Hebert. Recovering occlusion boundaries from a single image. In *ICCV*, 2007.

[9] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. In *CVPR*, 2008.

[10] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE PAMI*, 2006.

[11] V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts — A review. *IEEE PAMI*, 2007.

[12] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *IEEE PAMI*, 2004.

[13] N. Komodakis, N. Paragios, and G. Tziritas. MRF optimization via dual decomposition: Message-passing revisited. In *ICCV*, 2007.

[14] N. Komodakis and G. Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. In *CVPR*, 2006.

[15] S. Kumar and M. Hebert. A hierarchical field framework for unified context-based classification. In *ICCV*, 2005.

[16] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, 2001.

[17] X. Lan, S. Roth, D. Huttenlocher, and M. Black. Efficient belief propagation with learned higher-order markov random fields. In *ECCV*, 2006.

[18] V. Lempitsky, S. Roth, and C. Rother. FusionFlow: Discrete-continuous optimization for optical flow estimation. In *CVPR*, 2008.

[19] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[20] S. Ramalingam, P. Kohli, K. Alahari, and P. Torr. Exact inference in multi-label CRFs with higher order cliques. In *CVPR*, 2008.

[21] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *CVPR*, 2007.

[22] M. Sara and U. Sleytr. S-layer proteins. *Journal of Bacteriology*, 2000.

[23] J. Smit, H. Engelhardt, S. Volker, S. Smith, and W. Baumeister. The S-layer of caulobacter crescentus: 3d image reconstruction and structure analysisby electron microscopy. *Journal of Bacteriology*, 1992.

[24] D. Sontag, T. Meltzer, A. Globerson, T. Jaakkola, and Y. Weiss. Tightening LP relaxations for map using message passing. In *UAI*, 2008.

[25] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for MRFs with smoothness-based priors. *IEEE PAMI*, 2008.

[26] O. Veksler. Efficient graph-based energy minimization methods in computer vision, 1999.

[27] O. Veksler. Graph cut based optimization for MRFs with truncated convex priors. In *CVPR*, 2007.

[28] M. Wainwright, T. Jaakkola, and A. Willsky. MAP estimation via agreement on trees: message-passing and linear programming. *IEEE Trans. on Info. Theory*, 2005.

[29] C. Yanover, T. Meltzer, and Y. Weiss. Linear programming relaxations and belief propagation—an empirical study. *JMLR*, 2006.

[30] C. Yanover, O. Schueler-Furman, and Y. Weiss. Minimizing and learning energy functions for side-chain prediction. In *RECOMB*, 2007.