A Tiered Move-making Algorithm for General Pairwise MRFs

Vibhav Vineet Jonathan Warrell Philip H.S. Torr
Oxford Brookes University
{vibhav.vineet-2010, jwarrell, philiptorr}@brookes.ac.uk

Abstract

A large number of problems in computer vision can be modeled as energy minimization problems in a markov random field (MRF) framework. Many methods have been developed over the years for efficient inference, especially in pairwise MRFs. In general there is a trade-off between the complexity/efficiency of the algorithm and its convergence properties, with certain problems requiring more complex inference to handle general pairwise potentials. Graph-cuts based α-expansion performs well on certain classes of energies, and sequential tree reweighted message passing (TRWS) and loopy belief propagation (LBP) can be used for non-submodular cases. These methods though suffer from poor convergence and often oscillate between solutions.

In this paper, we propose a tiered move making algorithm *which is an iterative method. Each move to the next configuration is based on the current labeling and an optimal tiered move, where each tiered move requires one application of the dynamic programming based tiered labeling method introduced in Felzenszwalb et. al. [2]. The algorithm converges to a local minimum for any general pairwise potential, and we give a theoretical analysis of the properties of the algorithm, characterizing the situations in which we can expect good performance. We evaluate the algorithm on many benchmark labeling problems such as stereo, image segmentation, image stitching and image denoising, as well as random energy minimization. Our method consistently gets better energy values than α-expansion, LBP, quadratic pseudo-boolean optimization (QPBO), and is competitive with TRWS.*

1. Introduction

A large number of problems in computer vision can be modeled as discrete labeling problems. Examples include depth estimation from stereo, object segmentation and scene understanding, photomontage and image denoising [1, 2, 5, 7]. Random fields provide a general framework for building models for such problems, and can capture a wide range of phenomena which are relevant in many cases, such as smoothness and local interactions between labels.

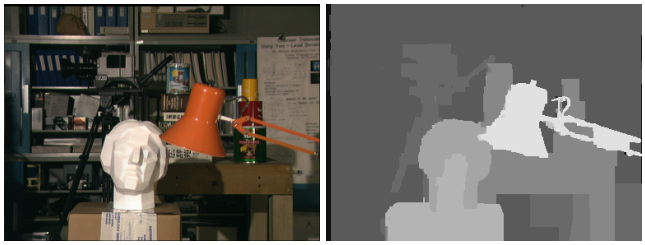

Figure 1. Our tiered move making algorithm achieves a minimum energy value 0.3% lower that that achieved by α-expansion on truncated L1-norm pairwise potential for stereo on the tsukuba image from [12] with 16 disparity labels. Further, we achieve around 0.01%, 0.1% and 1% lower energy values with Potts, non-truncated L1 and L2-norm pairwise potentials respectively. See experimental section for more details.

In general, such problems are NP-hard. Certain subclasses though are known to have polynomial time algorithms. Examples of these include binary labeling problems with submodular pairwise interaction terms [7], and multiple label problems with convex pairwise interactions [4]. The former can include for example foreground/background segmentation problems, while the latter can naturally be applied to cases in which an ordering exists across the labels, as in stereo. In both cases, the problem can be solved by building a graph to represent the problem, and solving an *s-t min cut* problem on the graph.

Constraining all problems to take such forms however is restrictive, and can lead to poor models in general. Examples of problems that cannot naturally be cast in such ways include multi-label problems without a label ordering (e.g. object class segmentation [5]), problems with an ordered label set but a truncated pairwise cost (e.g. state-of-the-art stereo models [7]), and binary problems with non-submodular pairwise terms.

A number of methods have been introduced which can be applied to solve these problems approximately. Approaches include belief propagation [15], linear programming relaxations [6, 8] and move-making algorithms [1, 3, 10, 11]. We will focus on the last of these approaches, although we note that connections have recently been drawn between graph-

cut based α-expansion moves and linear programming [9], as well as graph-cuts and belief propagation [13], pointing to a fundamental unity of these methods.

Move-making algorithms are designed to handle multi-label problems. They proceed by maintaining a *current solution* to the problem, and updating this in a series of moves, in which all or a subset of the pixels are allowed either to maintain their current label, or switch to a possibly restricted subset of labels. Moves are made only if they decrease the global energy of the random field, and termination occurs when no further move can decrease the energy [3]. Key factors which determine the effectiveness of a move-making algorithm are the suitability of the *move space* for the problem at hand, and the guarantees that can be made of the algorithm used to search this space. The α-expansion algorithm [1] for instance is typically effective on problems where large contiguous regions sharing the same label can be expected in the solution (e.g. segmentation problems), as the move space allows groups of pixels to switch to a common label in one move, and the optimal move can be found at each step for metric pairwise terms, which help to enforce piecewise smoothness. In other cases, acceptance of suboptimal moves can be compensated for by especially designed move-spaces. For instance, high-quality independently generated solutions can be exploited when available to propose alternative labels at each pixel as in *fusion moves* [11, 14], where QPBO can be used as the move generator [7]. Alternatively, the fact that large regions of the solution are expected to use restricted ranges of labels from an ordered set (e.g. stereo) can be exploited by using moves fix a range, and allow each pixel to swap to a label from that range as in *range expansion moves* [10], where the move generator is derived from [4].

From the approaches discussed above, while [1] uses exact (optimal) moves, it places metric conditions on the pairwise terms, and while [10, 11] can be applied to more general energies, [11] relies on independently generated solutions, [10] can only be applied to energies with truncated convex pairwise terms, and neither uses optimal moves. The question arises then whether a move-making approach can be proposed using an exact algorithm to generate optimal moves for general multi-label energies.[1] Here, we propose such an algorithm based on the dynamic programming approach to tiered labeling problems introduced in [2]. Our move space is thus defined by the tiered labeling constraints: at each move, a band of pixels across the image may change their labels, and each column of the image may choose a different label to change to. However, since multiple moves are made, the overall solution can be an arbitrary labeling. As we demonstrate, such moves are effective for a wide range of problems, for instance in which large contiguous label regions with convex shape are expected (segmentation of certain object classes) or many piecewise constant vertical regions are expected (stereo, see Fig. 1). We also show that the algorithm remains effective on problems where these assumptions do not obviously hold (photomontage and denoising). We compare against a range of other approximate algorithms, including α-expansion, loopy belief propagation (LBP), quadratic pseudo-boolean optimization (QPBO) and sequential tree reweighted message passing (TRWS), and show our algorithm consistently to achieve lower energies than all except for TRWS, with which we remain competitive. Finally, the fact we use optimal moves allows us to guarantee a bound on the quality of the global solution for certain classes of energy, which we show (under certain conditions) reduces to the α-expansion bound when a metric pairwise term is used.

In Sec. 1.1 we review the tiered labeling algorithm of [2]. We then show how this can be used as the basis of a move making algorithm on arbitrary energies in Sec. 2. Sec. 3 provides a theoretical analysis of the algorithm, characterizing its optimality and complexity properties, Sec. 4 gives an experimental comparison with other approaches, and Sec. 5 concludes with a discussion.

1.1. Tiered Labeling

We review here Felzenszwalb et al's *tiered labeling* algorithm [2], which we will use to form the basis of our move-making algorithm. The tiered labeling algorithm performs MAP inference in a pairwise MRF/CRF with general forms for the potentials subject to certain constraints on the final label configurations.

We let $p \in P$ range over the pixels of an image, and let $\mathbf{f} \in \hat{L}^{|P|}$ represent a labeling of the image, where $\hat{L}$ is a discrete label set, and we write $f_p = l$ to denote that pixel p is assigned label l. A pairwise MRF/CRF can be defined by an energy model, $E : f \rightarrow R$:

$$E(f) = \sum_{p \in P} D_p(f_p) + \sum_{pq \in N} V_{pq}(f_p, f_q) \qquad (1)$$

where $D_p(f_p)$ and $V_{pq}(f_p, f_q)$ are the unary and pairwise term respectively, and N is the set of neighbors for each pixel p in the image. For the tiered labeling algorithm, we assume there are two special members of $\hat{L}$, which we label T and B, for top and bottom. Writing $r_p \in \{1...m\}$ and $k_p \in \{1...n\}$ respectively for the row and column of pixel p in the image, where m and n are the number of rows and columns, we can express the constraints on the allowable tiered labeling solutions as:

[1] We note that [3] analyze γ-expansion moves, which are optimal in a stricter sense than we consider here. Our tiered moves make an optimal search over a fixed large space of transformations (Sec. 2).

$$\begin{aligned}\forall k \le n\ \exists i,j \le m, l \in \hat{L}\backslash\{T,B\}\text{ s.t.}\\ k_p = k \Rightarrow (r_p < i \ \Rightarrow\ f_p = T \wedge\\ r_p \ge j \ \Rightarrow\ f_p = B \wedge\\ i \le r_p < j \ \Rightarrow\ f_p = l)\end{aligned} \tag{2}$$

That is, we are allowed only two breakpoints i and j on each column, and above i the column is labeled T, below (and including) j it is labeled B, and between the two it takes a single label l from the remaining labels (where l can vary with the column). When $i = j$ label l is not used, and only T and B appear on the column.

Felzenszwalb et al. are able to minimize Eq. 1 subject to Eq. 2 exactly using dynamic programming. To do this, they first reexpress Eq. 1 as:

$$E(f) = \sum_{k=1}^{n} U_k(s_k) + \sum_{k=1}^{n-1} H_k(s_k, s_{k+1}) \tag{3}$$

Here, k ranges over the n columns of the image. Instead of expressing the energy in terms of pixel label assignments, f_p, Eq. 3 is expressed in terms of the variables $s_{k=1...n}$. These variables have the structure $s_k = (i_k,\ j_k,\ l_k)$, where i_k and j_k represent the two break-points for column k, as described above, and $l_k \in \hat{L}\backslash\{T,B\}$ is the label to be assigned to the pixels of the column for which $i_k \le r_p < j_k$. Any collection $\{s_1, s_2, ...s_n\}$ thus defines a labeling $\mathbf{f}$. The terms U and H represent unary and pairwise terms in this reformulated energy, and can be defined in terms of the previous energy terms by letting $P_k = \{p|k_p = k\}$, $N_k = \{(p,q)|(p,q) \in N, k_p = k_q = k\}$, $N_{k_1,k_2} = \{(p,q)|(p,q) \in N, k_p = k_1, k_q = k_2\}$, and $f_p(s_k)$ be the labeling of pixel $p \in P_k$ implied by s_k as:

$$\begin{aligned}U_k(s_k) &= \sum_{p \in P_k} D_p(f_p(s_k)) + \\ &\quad \sum_{pq \in N_k} V_{pq}(f_p(s_k), f_q(s_k))\\ H_k(s_k, s_{k+1}) &= \sum_{pq \in N_{k,k+1}} V_{pq}(f_p(s_k), f_q(s_{k+1}))\end{aligned} \tag{4}$$

Given these definitions, it is clear that minimizing Eq. 3 with respect to $s_{1...n}$ is equivalent to minimizing Eq. 1 with respect to f, subject to the constraints in Eq. 2.

Since Eq. 3 defines an energy on a chain, it can be solved via dynamic programming (DP). Writing n and m for the number of columns and rows respectively, and K for the number of labels (disregarding T and B), a naive DP implementation would have time complexity $O(nm^4K^2)$, since the number of states for variable s_k is m^2K. However, [2] introduce two tricks to reduce the time complexity to $O(nm^2K^2)$ by utilizing the special grid structure of MRF. The first involves the search for the best (i', j') in column $k-1$ given (i, j) in the column k, which would require $O(m^2)$ complexity for a brute-force search, but can be achieved in $O(m)$ time through a decoupling strategy. Second, the calculation of $U_k(s_k)$ and $H_k(s_k, s_{k-1})$ can be computed in $O(1)$ time using cumulative sums of the U, and H respectively. Further details on how the reduction of complexity is achieved can be found in [2].

2. Tiered move-making algorithm

In general, a move making algorithm can be expressed in terms of a *space of moves*, $t \in \mathcal{T}$, and a *transformation function* $\mathbb{T}$, which takes as input a labeling f^1 and transformation t, and outputs a second labeling f^2 (i.e. the result of applying t to f^1): $\mathbb{T}(f^1, t) = f^2$. The algorithm proceeds by generating a series of labelings across iterations $\nu = 1, 2, ...I$, such that either $E(f^{\nu+1}) = \text{argmin}_{t \in \mathcal{T}} E(\mathbb{T}(f^\nu, t))$, or $E(f^{\nu+1}) = \text{argmin}_{t \in \mathcal{T}'} E(\mathbb{T}(f^\nu, t))$, where $\mathcal{T}' \subset \mathcal{T}$. For some move spaces, the minimization cannot be made exactly, in which case a move t is chosen such that $E(f^{\nu+1}) = E(\mathbb{T}(f^\nu, t)) \le E(f^\nu)$. The algorithm terminates when $E(f^{\nu+1}) \ge E(f^\nu)$.

For our tiered move-making algorithm, we shall assume that we are minimizing a discrete energy function of the form Eq. 1, where $f_p = l \in L$ is a member of a generic label set L with no special members (such as T and B in Sec. 1.1) and subject to no constraints. We then consider the following move space, expressed in terms of vectors $\mathbf{t}$ over the augmented space of labels $\hat{L} = L \cup \{T, B\}$, i.e. $\mathbf{t} \in \hat{L}^P$. We shall require that these moves obey similar tiered constraints to those of Eq. 2:

$$\begin{aligned}\forall k \le n\ \exists i,j \le m, l \in \hat{L}\backslash\{T,B\}\text{ s.t.}\\ k_p = k \Rightarrow (r_p < i \ \Rightarrow\ t_p = T \wedge\\ r_p \ge j \ \Rightarrow\ t_p = B \wedge\\ i \le r_p < j \ \Rightarrow\ t_p = l)\end{aligned} \tag{5}$$

Our move space $\mathcal{T}$ is thus the subset of $\hat{L}^P$ satisfying constraints 5. Given a move $\mathbf{t}$, we can thus write i_k and j_k for the break points on column k, and $l_k \in L$ for the single label taken by t_p for p on column k and rows $i...j-1$. If $i = j$ we can arbitrarily set $l_k = 0$. Given a move $\mathbf{t}$, the transformation function for our move then switches all pixels p for which $t_p \neq \{T, B\}$ to the value t_p, and leaves all other variables unchanged. Writing $\mathbb{T}_p(f, \mathbf{t})$ for the value of $\mathbb{T}(f, \mathbf{t})$ at pixel p, we can thus express the transformation as:

$$\mathbb{T}_p(f, \mathbf{t}) = \left\{ \begin{array}{cc} t_p & \text{if } t_p \neq \{T, B\} \\ f_p & \text{otherwise} \end{array} \right\} \tag{6}$$

As described above for the general case, we generate a sequence of labelings $f^{\nu=1...I}$, which is initialized randomly, and satisfies $E(f^1) > E(f^2) ... > E(f^I)$. Using

Algorithm 1: Tiered move making algorithm

input : Energy function E, initial labeling f^1
converged := *0*, ν := *1*;
while *converged* $= 0$ **do**
  Calculate D^ν and V^ν for $E^\nu(\mathbf{t})$ (Eq. 8);
  Calculate U and H terms (Eq. 4);
  Solve tiered labeling problem for $\mathbf{t}^* = \text{argmin}_{\mathbf{t}}\, E^\nu(\mathbf{t})$;
  Construct $f^{\nu+1}$ from $\mathbf{t}^*$ (Eq. 10);
  if $E(f^{\nu+1}) < E(f^\nu)$ **then**
    $\nu := \nu + 1$;
  else
    converged = *1*;
  end
end
Return f^ν;

the tiered labeling algorithm of Sec. 1.1, we can perform an optimal search at each iteration, so that $E(f^{\nu+1}) = \text{argmin}_{\mathbf{t}\in\mathcal{T}}\, E(\mathbb{T}(f^\nu, \mathbf{t}))$. We achieve this by transforming the search for the optimal $\mathbf{t}$ into a tiered labeling problem as follows. We can reexpress the energy of a particular transformation $E(\mathbb{T}(f^\nu, \mathbf{t}))$ directly as an energy over the transformation vectors $\mathbf{t}$ as:

$$E^\nu(\mathbf{t}) = \sum_{p\in P} D_p^\nu(t_p) + \sum_{pq\in N} V_{pq}^\nu(t_p, t_q) \tag{7}$$

where the data and pairwise terms in Eq. 7 are defined as:

$$D_p^\nu(t_p) \quad = \quad \left\{ \begin{array}{cc} D_p(t_p) & \text{if } t_p \in L \\ D_p(f_p^\nu) & \text{otherwise} \end{array} \right\}$$

$$V_{pq}^\nu(t_p, t_q) = \left\{ \begin{array}{cc} V_p(t_p, t_q) & t_p, t_q \in L \\ V_p(t_p, f_q^\nu) & t_p \in L, t_q \in \{T, B\} \\ V_p(f_p^\nu, t_q) & t_p \in \{T, B\}, t_q \in L \\ V_p(f_p^\nu, f_q^\nu) & \text{otherwise} \end{array} \right\} \tag{8}$$

It is clear then that solving the minimization:

$$f^{\nu+1} = \text{argmin}_{\mathbf{t}\in\mathcal{T}}\, E(\mathbb{T}(f^\nu, \mathbf{t})) \tag{9}$$

is equivalent to minimizing Eq. 7, subject to the constraints Eq. 5, which can thus be achieved via the tiered labeling approach of Sec. 1.1. Having found $\mathbf{t}^* = \text{argmin}_{\mathbf{t}}\, E^\nu(\mathbf{t})$, we can then generate $f^{\nu+1}$ via:

$$f_p^{\nu+1} \quad = \quad \left\{ \begin{array}{cc} t_p^* & \text{if } t_p^* \in L \\ f_p^\nu & \text{otherwise} \end{array} \right\} \tag{10}$$

As described above, we terminate when $E(f^{\nu+1}) = E(f^\nu)$. The algorithm is summarized in Alg. 1.

3. Algorithm properties and analysis

Since our tiered move making algorithm is based on dynamic programming, it can be applied to arbitrary energy pairwise functions, and is guaranteed to converge, since $E(f^\nu) > E(f^{\nu+1})$ at each iteration, and given a finite label set L there are only finitely many labelings. If we make further assumptions about the energy, we can also give certain guarantees on the strength of the local optimum our algorithm achieves, as described in Sec. 3.1. Further, we discuss the complexity of the algorithm in Sec. 3.2.

3.1. Optimality

We show here that we can give guarantees about the local optimum achieved by our algorithm if we restrict the class of energy functions we consider to ones in which $\forall p, l\ \ D_p(l) \geq 0$, and $\forall p, q, l_1, l_2\ \ V_{pq}(l_1, l_2) \geq 0$, with equality occurring if and only if $l_1 = l_2$. We note that the assumption that the unary terms are non-negative is not restrictive, as we can always add a constant a to all outputs of $D_p(.)$ to create a new potential $D'_p(l) = D_p(l) + a$, and replacing $D_p(.)$ by $D'_p(.)$ simply adds a constant to the global energy $E'(f) = E(f) + a$, i.e. the ordering of solutions by energy remains unchanged. The restriction on the pairwise terms covers many cases of interest, such as semi-metric potentials (which additionally require $\forall p, q, l_1, l_2\ \ V_{pq}(l_1, l_2) = V_{pq}(l_2, l_1)$), metric potentials (which add to the semimetric requirements $\forall p, q, l_1, l_2, l_3\ \ V_{pq}(l_1, l_2) + V_{pq}(l_2, l_3) \geq V_{pq}(l_1, l_3)$), and pairwise Potts potentials (which require $\forall p, q\ \ \exists v\ \ \forall l_1 \neq l_2\ \ V_{pq}(l_1, l_2) = v$). Additionally, we are interested in labelings of the image that are *tiered consistent*; that is, given a labeling f, and an associated set of *maximal contiguous regions* R^f whose members must take a single label, be connected, and have non-matching neighboring pairs on their boundary (i.e. $R^f = \{\rho \subset P | (\exists l \in L,\ \forall p \in \rho\ (f_p = l))$ and $(\forall p, q \in \rho,\ \exists Z < P,\ (p_1, p_2, ...p_Z) \in \rho^Z\ (p = p_1 \wedge q = p_Z \wedge (\forall z < Z\ ((p_z, p_{z+1}) \in N))))$ and $(\forall p, q \in P\ ((p, q) \in N \wedge p \in \rho \wedge q \notin \rho \Rightarrow f_p \neq f_q))\}$), we can define the set of tiered consistent labelings C as:

$$\begin{aligned} C \quad = \quad & \{f | \forall \rho \in R^f, k \leq n\ \exists i, j\ s.t. \\ & k_p = n \wedge (i \leq r_p < j) \Rightarrow p \in \rho\} \end{aligned} \tag{11}$$

Fig. 2 illustrates the properties introduced above. Given these definitions, we can make the following guarantee:

Theorem 1. *Let $\hat{f}$ be a local minimum for tiered labeling moves and $f^\dagger = \text{argmin}_{f\in C} E(f)$ be the optimal tiered consistent labeling. Then for energies with non-negative unary terms and in which $\forall p, q, l_1, l_2\ V_{pq}(l_1, l_2) \geq 0$, with equality occurring if and only if $l_1 = l_2$, we have that $E(\hat{f}) \leq 2\kappa E(f^\dagger)$, where $\kappa = \max_{p,q\in N}(\max_{l_1\neq l_2\in L} V_{pq}(l_1, l_2) / \min_{l_1\neq l_2\in L} V_{pq}(l_1, l_2))$.*

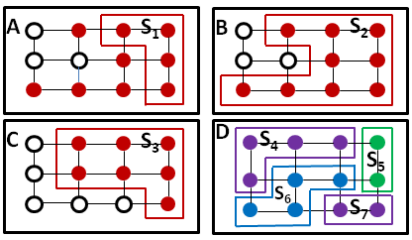

Figure 2. Maximal contiguous regions and tiered consistency. Colours correspond to different labels. (A) S_1 is not a *maximally contiguous region*, since there are neighbour pairs on its boundary sharing a common label. (B) S_2 is maximally contiguous but not *tiered consistent*, since its intersection with column 2 is disconnected (requiring 4 break-points to represent). (C) S_3 is both maximally contiguous and tiered consistent. (D) In a *tiered consistent labeling* ($f \in C$) all maximally contiguous regions ($S_4, S_5, S_6, S_7 \in R^f$) are tiered consistent. Note that a single label may give rise to multiple maximally contiguous regions, e.g. S_4 and S_7. See Sec. 3.1.

Proof: We consider the set of maximally contiguous regions $\rho \in R^{f^\dagger}$. Given any ρ, we define f^ρ to be:

$$f^\rho = \left\{ \begin{array}{cc} f^\dagger_p & \text{if } p \in \rho \\ \hat{f}_p & \text{otherwise} \end{array} \right\} \tag{12}$$

We can assume $E(\hat{f}) \geq E(f^\dagger)$, since otherwise the bound follows directly. Since f^ρ is within a tiered move of $\hat{f}$, we have:

$$E(f^\dagger) \leq E(\hat{f}) \leq E(f^\rho) \tag{13}$$

Further, we can define the sets $I^\rho = \rho \cup \{(p,q) \in N | p,q \in \rho\}$, $B^\rho = \{(p,q) \in N | p \in \rho, q \notin \rho\}$, and $O^\rho = (P \backslash \rho) \cup \{(p,q) \in N | p,q \notin \rho\}$, which represent respectively, all pixels and neighbour pairs falling inside ρ (I^ρ), neighbour pairs between ρ and $P \backslash \rho$ (B^ρ), and all pixels and neighbours outside of ρ(O^ρ). We then have:

$$\begin{array}{rcl} E(f^\rho | O^\rho) & = & E(\hat{f} | O^\rho) \\ E(f^\rho | I^\rho) & = & E(f^\dagger | I^\rho) \\ E(f^\rho | B^\rho) & \leq & \kappa E(f^\dagger | B^\rho) \end{array} \tag{14}$$

where we write $E(f|O^\rho)$ for the value of the energy Eq. 1 restricted to potentials on pixels/neighbours in O^ρ, and similarly for $E(f|I^\rho)$, $E(f|B^\rho)$. The constant κ is defined as stated in the Theorem, $\kappa = \max_{p,q \in N} (\max_{l_1 \neq l_2 \in L} V_{pq}(l_1,l_2) / \min_{l_1 \neq l_2 \in L} V_{pq}(l_1,l_2))$, and we can guarantee the last of these inequalities, since the requirement that ρ is a *maximal* contiguous set implies that $\forall (p,q) \in B^\rho \; f^\dagger_p \neq f^\dagger_q$, and the requirements on the energy imply therefore $\forall (p,q) \in B^\rho \; V_{pq}(f^\dagger_p, f^\dagger_q) > 0$. Now, by separating the final inequality in Eq. 13, we get the expansion $E(\hat{f}|I^\rho) + E(\hat{f}|B^\rho) + E(\hat{f}|O^\rho) \leq E(f^\rho|I^\rho) + E(f^\rho|B^\rho) + E(f^\rho|O^\rho)$, and substituting in the values from Eq. 14 we get $E(\hat{f}|I^\rho) + E(\hat{f}|B^\rho) + E(\hat{f}|O^\rho) \leq E(f^\dagger|I^\rho) + \kappa E(f^\dagger|B^\rho) + E(\hat{f}|O^\rho)$, which reduces to $E(\hat{f}|I^\rho) + E(\hat{f}|B^\rho) \leq E(f^\dagger|I^\rho) + \kappa E(f^\dagger|B^\rho)$. To bound the whole energy, we must sum this final inequality over all $\rho \in R^{f^\dagger}$, giving $\sum_{\rho \in R^{f^\dagger}} (E(\hat{f}|I^\rho) + E(\hat{f}|B^\rho)) \leq \sum_{\rho \in R^{f^\dagger}} (E(f^\dagger|I^\rho) + \kappa E(f^\dagger|B^\rho))$. Letting $B = \bigcup_{\rho \in R^{f^\dagger}} B^\rho$ and collecting together terms, we can reexpress this as:

$$E(\hat{f}) + E(\hat{f}|B) \leq E(f^\dagger) + (2\kappa - 1) E(f^\dagger|B) \tag{15}$$

and subtracting $E(\hat{f}|B) \geq 0$ from the left hand side of Eq. 15 and adding $(2\kappa - 1)(E(f^\dagger) - E(f^\dagger|B)) \geq 0$ to the right we have the final bound $E(\hat{f}) \leq 2\kappa E(f^\dagger)$. □

Corollary. *For energies of the form in Theorem 1, and for which the optimal labeling* $f^* = \operatorname{argmin}_f E(f)$ *is tiered consistent, we have that* $f^* = f^\dagger$*, and thus we achieve the global bound* $E(\hat{f}) \leq 2\kappa E(f^*)$.

For cases in which the optimal energy f^* is tiered consistent, we can thus achieve the same bound as α-expansion [1] for energies with metric and Potts pairwise potentials. We note that, when $f^\dagger = f^*$, this bound actually holds for energies with a larger class of pairwise potentials as specified by Theorem 1, including semi-metrics and a restricted class of non-symmetric potentials. In fact, for the pairwise case these are the same restrictions on the energy as in a recent generalization of the α-expansion bound in [3]. Our bound is thus identical to [3] for energies with a tiered consistent f^*, although we cannot apply the [3] bound directly to the analyze tiered moves, since they do not form an optimal γ-expansion move in the sense required by [3].

Clearly, we cannot know for any given energy whether or not $f^\dagger = f^*$. For many problems though, such as segmentation of objects with approximately convex shapes, or stereo, where we can expect regions of constant disparity to be approximately vertical, we can expect that this condition will hold approximately, or equivalently that the violations of the tiered constraints in f^* will be small (assuming the energy itself provides a good representation of the problem). This argument lends theoretical support to our claim in Sec. 4 that tiered moves are a competitive alternative to α-expansion moves (and non-metric variants) for a wide range of problems.

Worst case suboptimality: Although we have argued above that for many problems of interest, we might expect $f^\dagger \approx f^*$, we now construct a simple binary example to show that in general the energy of $f^\dagger$ can be arbitrarily far from the optimum. Consider for example a binary energy over $P = 6$ pixels with $m = 3$ rows and $n = 2$ columns, with column-major indexing (see Fig. 3). Assuming a 4-connected graph, and Potts pairwise terms between neighbours, we define an energy such that

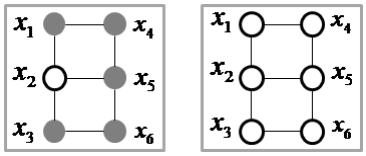

Figure 3. Suboptimal binary case: Left shows the optimal configuration f^* which is non-tiered consistent ($E(f^*) = 1$), but right shows the best tiered consistent labeling $f^\dagger$ having suboptimal energy ($E(f^\dagger) = Q$). White indicates variables with value 0, and black indicates values 1. See Sec. 3.1 (Worst case suboptimality).

$D_1(0) = D_3(0) = Q/2$ and $D_2(1) = 2Q$ ($Q > 1$), while the non-zero term for pairwise potentials $V_{1,2}$,$V_{2,3}$ and $V_{2,5}$ is set to $1/3$ and for $V_{1,4}$,$V_{4,5}$,$V_{5,6}$ and $V_{3,6}$ is set to $2Q$. Writing $f = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ for a joint setting of the 6 variables, we can determine that the global optimum is $f^* = \{1, 0, 1, 1, 1, 1\}$, where $E(f^*) = 1$. This labeling however is not tiered consistent, as it contains a maximal contiguous region of 1's with 4 break-points on the first column. The optimal tiered consistent configuration is $f^\dagger = \{0, 0, 0, 0, 0, 0\}$, where $E(f^\dagger) = Q$ (see Fig. 3). For all other labelings $E(f) \geq 2Q$. As well as being the optimal tiered consistent labeling, $f^\dagger$ is also a local optimum for tiered moves (i.e. it forms a possible $\hat{f}$), since the only configuration with lower energy cannot be reached in a single move (in general, the algorithm can converge to arbitrary labelings). We note that, by construction $E(f^\dagger) = Q$ can be increased without limit, and so can be arbitrarily far from the global optimum $E(f^*) = 1$. This example shows that our algorithm can fail when the global optimum contains closely connected non-convex regions.

3.2. Complexity

The complexity of the individual moves in our algorithm can be derived from the complexity of the tiered labeling algorithm itself, which dominates the move:

Theorem 2. *The time complexity of each individual tiered move is $O(nm^2K^2)$, where n, m are the number of image columns and rows respectively, and K is the number of labels in the global energy.*

Proof: We first note that application of the tiered labeling algorithm dominates the move, since computing D^ν, V^ν, U and H are all $O(nmK)$, and applying Eq. 10 to update the solution is $O(mn)$. As described in Sec. 2, finding the optimal move involves solving a tiered labeling problem with an expanded label set $\hat{L} = L \cup \{T, B\}$. Since labels T, B cannot appear in the dynamic programming variables $s_{k=1\ldots n} = (i_k, j_k, l_k)$ we have a complexity $O(nm^2K^2)$ for each move, following [2]. □

4. Experiments

In this section we present experimental results on several low-level benchmark problems, including stereo correspondence, image segmentation, image stitching and image denoising. The review paper of Szeliski et. al. [12] provides some benchmark images corresponding to each problem, some of which we rescale (see Fig. 4). We compare our tiered move-making algorithm (t-move) with four other energy minimization methods, α-expansion, QPBO, max-product loopy belief propagation (LBP), and sequential tree re-weighted message passing (TRWS) on these benchmark applications, and also the original tiered labeling algorithm [2] for segmentation. We use code provided by the authors for the QPBO method [7] and codes in the software library provided by Szeliski et. al. [12] for LBP, α-expansion and TRWS. We use QPBO to perform expansion moves over a series of iterations similar to α-expansion, as in Woodford et.al. [14]. The pairwise potentials we use for our comparative study are Potts models, truncated/non-truncated linear and quadratic functions. Further, to show the generalization ability of our algorithm, we also conduct experiments on random pairwise potentials.

Stereo: Given left and right rectified images of a scene, the stereo problem requires us to calculate the disparity of each pixel in the left image to a pixel in right image. Discrete labels are used to represent disparities in terms of pixel displacement. We conduct experiments on the tsukuba, venus, and teddy images; these three benchmark images are shown in Fig. 4(**A**). In all these examples, we use the absolute intensity difference between two corresponding pixels in the left and the right images as a unary potential.

Table 1 shows the comparative lowest energy values achieved by α-expansion, LBP, TRWS and tiered moves on the three benchmark images. The smoothness terms are a Potts model, and linear and quadratic pairwise potentials. We observe that our method consistently achieves lower energy values than α-expansion, QPBO, and LBP on all three images. We observe that it also achieves lower energy values than TRWS on the teddy image for the Potts model. Further, the lowest energy values achieved by our method are almost equal to the lower bound on the energy values provided by TRWS method. In all these cases, t-moves are able to achieve almost $1\% - 2\%$ lower energies than α-expansion, QPBO and LBP methods, though TRWS achieves slightly better energies than our method. We also compared our method against α-expansion with a truncated L1-norm and achieved $\sim 0.3\%$ lower energies.

Segmentation: We select the person, flower, and sponge images as shown in Fig. 4(**B**) for binary segmentation. Our data-term is based on gaussian mixture colour models of the foreground and background, and the smoothness term is a Potts model modulated by local contrast values. The corresponding energy values for all these images and benchmark algorithms are shown in the Table 2. We achieve the global minimum energy values on all images, as does α-expansion. Further, we achieve almost $2\% - 3\%$ lower energy values than the original tiered labeling method on all images.

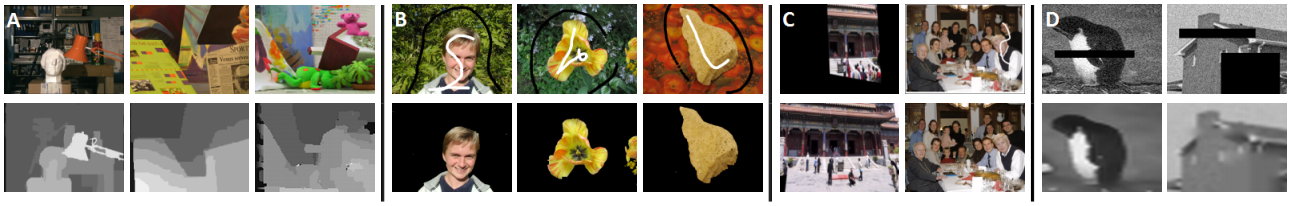

Figure 4. Shown are the input (first row) and output (second row) images of our tiered move making algorithm corresponding to different benchmark problems. (From left to right) **Stereo (A)**: tsukuba (192x144, 8 disparities), venus (217x192, 10 disparities) and teddy (112x93, 15 disparities) images; **Segmentation (B)**: person (600x450), flower (600x450) and sponge (640x480) images; **Image stitching (C)**: pano (178x120, 7 labels), and family (376x283, 5 labels); **Image denoising (D)**: penguin (30x44, 256 labels), and family (32x32, 256 labels).

Image	Method	Potts	L1-norm	L2-norm
Tsukuba	Lower Bound	50772	106407	98450
	LBP	52261	112659	100461
	TRWS	50779	106407	98451
	α-exp	50848	106529	104886
	QPBO	51471	113593	103823
	t-move	50803	106411	99723
Venus	Lower Bound	97270	160417	138682
	LBP	97634	163475	142966
	TRWS	97277	160417	138682
	α-exp	97463	160512	145627
	QPBO	98503	175850	145042
	t-move	97279	160436	138877
Teddy	Lower Bound	56032	126601	81084
	LBP	56641	128845	84057
	TRWS	56051	126601	81084
	α-exp	56093	126631	82898
	QPBO	56327	131607	83571
	t-move	56039	126659	81922

Table 1. Stereo: Compares the lowest energy values achieved by tiered moves and other benchmark energy minimization methods on three different stereo images for Potts, linear and quadratic pairwise potentials. The lower bound on the energy values given by TRWS is also shown.

Image	Method	Smooth	Data	Total
Person	t-label	682815	32890529	33573344
	α-exp	647809	30157043	30804852
	t-move	647809	30157043	30804852
Flower	t-label	636004	40235734	40871738
	α-exp	679744	39831530	40511274
	t-move	679744	39831530	40511274
Sponge	t-label	273505	31974735	32248240
	α-exp	255933	31471232	31727165
	t-move	255908	31471257	31727165

Table 2. Segmentation: We compare the lowest energy values achieved by different benchmark energy minimization methods on three different segmentation images. We achieve the global minimum in all cases, as does α-expansion. Results from the original tiered labeling algorithm (t-label) [2] are also given.

Image	Method	Smooth	Data	Total
Family (148956)	TRWS	149784	0	149784
	α-exp	149181	0	149181
	t-move	149024	0	149024
Pano (80813)	TRWS	80813	0	80813
	α-exp	80813	0	80813
	t-move	80813	0	80813

Table 3. Image Stitching: Compares lowest energy values achieved by different benchmark energy minimization methods on two different image stitching benchmarks. The lower bound on energy values given by TRWS is also shown in the first column.

Photomontage: Given a set of overlapping images, photomontage requires us to stitch them into one seamless image. The data-term is either 0 or infinity, depending on whether that pixel is in the field of view of the current camera direction. We show our results on the family and panorama images in Fig. 4(**C**). Our pairwise terms depend on pixel location and labels, as given in Szeliski et.al. [12]. In this case, our method achieves lower energy values than all the other benchmark methods shown in Table 3. We achieve almost $0.2\% - 0.5\%$ lower energy than TRWS and α-expansion on the family image set. For the pano image, we achieve the lower bound on the energy value (i.e. the global optimum) as achieved by all the other methods.

Image denoising: In the denoising problem, we are given a noisy image, and the task is to output a denoised image. This problem is also formulated as an energy minimization problem over an MRF where each pixel can receive any label from 256 intensity values. We conduct experiments on the house and penguin images in Fig. 4(**D**). Our data term is the squared difference between the actual pixel value and the output label, and the smoothness term is the squared difference between labels, weighted by a common factor $\lambda = 5$. In all these benchmark images, we achieve energies lower than α-expansion by almost 5%, although our energies are slightly higher than TRWS (see Table 4).

Random energies: In order to show the generalization ability of our algorithm in handling any pairwise potential, we compare the minimum energy values achieved by different benchmark methods on random energy functions.

Image	Method	Smooth	Data	Total
Penguine (1776156)	TRWS	270425	1505731	1776156
	α-exp	210075	1778962	1989037
	t-move	261150	1521812	1782962
House (605613)	TRWS	181285	424329	605614
	α-exp	202670	419748	622418
	t-move	179710	426549	606259

Table 4. Denoising: Compares lowest energy values achieved by different benchmark energy minimization methods on two different image denoising images. The lower bound on energy values given by TRWS is also shown in the first column.

We generate a set of random energy functions with random unary and pairwise potentials. We observe that the QPBO method consistently does quite well, followed by our tiered move method. The minimum energy values achieved by QPBO are generally within $(1-2)\%$ of the TRWS bound, for tiered-moves within $(4-5)\%$, for LBP within $(6-7)\%$, for α-expansion within $(10-11)\%$, and TRWS performs poorly in all the cases with minimum energy values lying within $(15-16)\%$ of the lower bound it generates.

5. Discussion and Conclusion

In this paper, we propose a tiered move making algorithm. Each iteration of algorithm makes a decision whether to retain the current labeling or to take a new label based on the optimal tiered move step. Each optimal tiered move is found by applying the dynamic programming based tiered labeling method.

Our method is guaranteed to converge to a local minima for any pairwise potential, which is not guaranteed by other inference methods. Graph cuts based α-expansion will not converge to a local minimum when functions are non-metric and non-submodular, since generally α-expansion solvers truncate the pairwise potential. Similarly, message passing based methods like LBP and TRWS are not guaranteed to converge and the final solution may oscillate between two different labelings. However, the TRWS method interestingly provides a lower bound on the energy, which provides a basis of comparison for other algorithms. Our tiered move method always converges to a local optimum, and does not suffer from the problem of oscillation.

The QPBO method [7] is a graph-cut based method which is able to handle non-submodular energy functions. But, this method does not always label all the nodes, which leads to partial labelings. The unlabeled pixels are labeled heuristically. However, we do not suffer from this problem since we do not allow fractional solutions.

Our tiered move making algorithm however suffers from certain limitations. The time complexity of this method is $O(nm^2K^2)$ per move, and in practice this method is almost 10-15 times slower than α-expansion. However, this move complexity could be reduced by using schemes such as γ-expansions [3] where only subsets of the labels are expanded at each move, without affecting the guarantees in Sec. 3.1. Another possibility is to implement the tiered labeling method on GPU, which would involve distributing the computation per state over threads [2]. Further, as with the tiered labeling method, the space complexity is $O(nm^2K)$ due to the dynamic programming requirements. Also, we cannot guarantee that a series of tiered moves will lead to the globally optimal solution even in the 2 label sub-modular case as is guaranteed by graph cuts (see Sec. 3.1). However, empirically we achieve the global minimum for many of the problems investigated, and we are mostly within 0.001% of the global minimum otherwise. We believe that only if there are many closely connected non-convex shapes in the solution will our method get stuck in bad local minima.

As our experiments suggest, we consistently do well compared to α-expansion, LBP, and QPBO and are very competitive with the TRWS method on a wide range of problems, including stereo, segmentation, image stitching, denoising, as well as optimization of random energies. Our theoretical analysis provides further support for the expectation that our algorithm will remain competitive in many commonly encountered problems, and the code is available for download at http://cms.brookes.ac.uk/staff/VibhavVineet/.

References

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

[2] P. F. Felzenszwalb and O. Veksler. Tiered scene labeling with dynamic programming. In *CVPR*, 2010. 1, 2, 3, 6, 7, 8

[3] S. Gould, F. Amat, and D. Koller. Alphabet soup: A framework for approximate energy minimization. In *CVPR*, 2009. 1, 2, 5, 8

[4] H. Ishikawa. Exact optimization for markov random fields. *PAMI*, 25(10), 2003. 1, 2

[5] P. Kohli, M. P. Kumar, and P. H. S. Torr. P3 & beyond: Move making algorithms for solving higher order functions. *PAMI*, 31(9), 2009. 1

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

[7] V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts-a review. 29(7), July 2007. 1, 2, 6, 8

[8] N. Komodakis, N. Paragios, and G. Tziritas. Mrf optimization via dual decomposition: Message-passing revisited. In *ICCV*, 2007. 1

[9] N. Komodakis and G. Tziritas. A new framework for approximate labeling via graph cuts. In *CVPR*, 2006. 2

[10] M. Kumar, O. Veksler, and P. Torr. Improved moves for truncated convex models. *JMLR*, 12, 2011. 1, 2

[11] V. Lempitsky, C. Rother, S. Roth, and A. Blake. Fusion moves for markov random field optimization. *PAMI*, 32(8), 2010. 1, 2

[12] 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 markov random fields with smoothness-based priors. *PAMI*, 30(6), 2008. 1, 6, 7

[13] D. Tarlow, I. Givoni, R. Zemel, and B. Frey. Graph cuts is a max-product algorithm. In *NIPS*, 2011. 2

[14] O. Woodford, I. Reid, P. Torr, and A. Fitzgibbon. Fields of experts for image-based rendering. In *BMVC*, volume 3, 2006. 2, 6

[15] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propagation. In *NIPS*, 2000. 1