Uncertainty Driven Multi-scale Energy Minimization

Pushmeet Kohli   Victor Lempitsky   Carsten Rother

Microsoft Research Technical Report, April 2010

Abstract

This paper proposes a new multi-scale energy minimization algorithm which can be used to efficiently solve large scale labelling problems in computer vision. The basic modus operandi of any multi-scale method involves the construction of a smaller problem which can be solved efficiently. The solution of this problem is used to obtain a partial labelling of the original energy function, which in turn allows us to minimize it by solving its (much smaller) projection. We propose the use of new techniques for both the construction of the smaller problem, and the extraction of a partial solution. We demonstrate our method on the problem of interactive image segmentation. Traditional multi-scale approaches for segmentation extract a partial solution using an image band around the boundaries of the object segmentation obtained by minimizing the smaller problem. This strategy fails on objects with fine structures and complex topologies. In contrast, our novel approach uses a min-marginal based uncertainty measure which allows us to handle such objects. Experiments show that our techniques result in solutions with low pixel labelling error. Furthermore, they take the same or less amount of computation compared to traditional multi-scale techniques.

1. Introduction

Energy minimization and discrete optimization have become a cornerstone of computer vision. This has primarily been driven by their ability to efficiently compute the Maximum a Posteriori (MAP) solutions in models such as Markov and Conditional random fields (MRFs and CRFs) [3, 5, 10, 11]. For instance, the widely used maximum flow algorithms allow for the efficient computation of the globally optimal solutions of certain cost (energy) functions, and have enabled the development of interactive image and 3D volume segmentation tools.

In recent years, advances in image acquisition technologies have significantly increased the size of images and 3D volumes. For instance, the latest commercially available cameras can capture images with almost 20 million pixels. In fact it is now possible to capture giga-pixel images of complete cities [17]. Similarly, latest medical imaging systems can acquire 3D volumes with billions of voxels. This type of data gives rise to large scale optimization problems which are extremely computationally expensive to solve and require large amounts of memory. Due to these factors, the development of efficient and scalable energy minimization algorithms has become a very active field of research and a number of multi-scale [18, 25, 27] and distributed algorithms [7] have been proposed for such problems.

Multi-Scale Methods and Related Work   Multi-scale processing has long been a popular approach to reduce the memory and computational requirements of optimization algorithms (see [21] for a review). It has been successfully used for efficiently solving many different labelling problems in computer vision [16]. The basic structure of these methods is quite simple. In order to label a large image (or 3D volume) they first solve the problem at low resolution, obtaining a coarse labelling of the original high resolution problem. This labelling is refined by solving another optimization on a small subset of the pixels.

A classic example of such a multi-scale method is the boundary band algorithm of Lombaert et al. [19], which addressed the problem of segmenting large images and 3D volumes. The algorithm worked by constructing low resolution versions of the image (or voxel volume) to be segmented. It first processes the low resolution image resolution to obtain a coarse segmentation labelling (figure 1c). A narrow band surrounding the foreground/background boundary (figure 1d) is used to isolate the image regions that need to be analyzed at the finer resolution. The final segmentation solution (figure 1b) is obtained by minimizing an energy function which depends only on the variables isolated in the narrow band. This operation requires dramatically less computation resources compared to the problem of minimizing the energy functions with all variables.

The algorithm described above suffers form the problem that it cannot efficiently recover from large errors present in the coarse labelling. For instance, if a thin foreground structure is missed in the coarse labelling, a large band of pixels will need to be analyzed at the fine scale. This would make the size of the resulting higher resolution problem large and reduce the computational benefits. An interest-

ing method to resolve this problem was proposed Sinop and Grady [26]. Motivated by the problem of segmenting thin structured objects they used the information from a Laplacian pyramid [6] to isolate pixels which might not have attained their correct labelling at the low resolution image.

Recently, Lempitsky and Boykov [18] presented an interesting *touch-expand* algorithm that is able to minimize pseudo-boolean energy function using a narrow band, while retaining the global optimality guarantees. On the downside, it has no bounds on the size of the band it may need to consider, and in the worst case the band can progressively grow to encompass the whole image. While for highly structured unary terms concerned with the shape fitting task considered in [18] the bands are reasonably small, they are highly likely not to be so for the less structured unary terms in e.g. segmentation problems.

1.1. Our Contributions

The goal of this paper is to develop a multi-scale algorithm which can be used minimize energy function with a large number of variables. To do this, we need to answer the following questions: **(1)** How to construct the energy for the small scale problem? **(2)** After minimizing this energy, how do we then isolate variables which need to be solved at the finer resolution? As we will explain later in the paper, the answers to these two important questions are not independent. We will now provide a brief overview of our strategy[1]. For ease of explanation, we will use the two label interactive image segmentation problem as an example. However, our method is general and can be used for any labelling problem such as 3D reconstruction, stereo, object segmentation and optical flow.

Constructing the Low Resolution Energy Construction of the energy defined in the small coarse grid is a crucial part of any multi-scale segmentation method. Ideally, we would want to construct the energy function in such a manner that its optimal solution, when projected to the full grid, matches the optimal solution of original energy as closely as possible. Recent band-based methods for image segmentation such as [19] and [26] construct the small scale energy from a low-resolution version of the image to be segmented. In contrast, our approach constructs the small scale energy directly from the energy of the full resolution image. Experiments show that this strategy results in substantial improvements in running time and accuracy.

Uncertainty Driven Bands The traditional band-based multi-scale segmentation methods use the MAP (or lowest energy) solution of the small scale problem to isolate which pixels need to be solved at the fine scale. These techniques
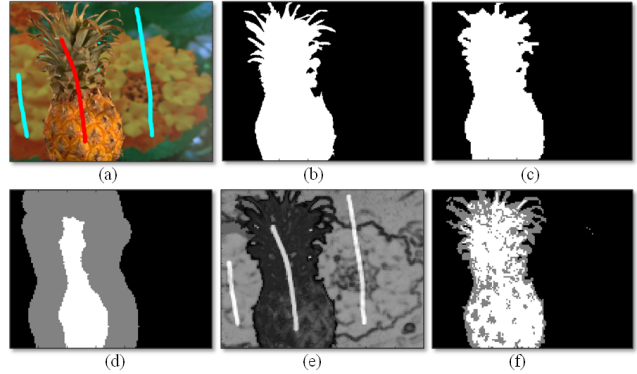
---

[1]Exact details will come later in the paper.

Figure 1. *Uncertainty driven multi-scale image segmentation.* **(a)** *Image with user marked brush strokes for different segments: foreground (red) and background (blue).* **(b)** *Segmentation obtained by minimizing the conventional segmentation energy (2) defined over the image grid. White pixels represent foreground, while black pixels represent background.* **(c)** *Segmentation obtained by minimizing an energy defined over a coarse level gird which is constructed using our method (see section 4.4). Observe that many pixels take labels different from the MAP labels shown in (b). To correct such errors we need to mark such pixels as unlabelled, and find their labels at the fine scale.* **(d)** *Partial segmentation obtained by marking pixels in the band around the segmentation boundary of (c) as unlabelled (marked gray). The size of the band is choosen to include all incorrectly labelled pixels.* **(e)** *Min-marginal based confidence values for pixels taking the MAP label (bright pixels are more confident). (see section 5.2 for more details).* **(f)** *Partial labelling obtained by marking pixels below a confidence score as unlabelled. As for (d) the confidence threshold is choosen to include all incorrect pixels. For this example, with the uncertainty based scheme 3 times less number of pixels need to marked as unlabelled compared to the number marked with the boundary band. As less number of variable need to be solved at the fine resolution, we get a bigger speedup.*

do not take into account the confidence or uncertainty associated with the MAP label assignment for any variable. Intuitively, if a variable has low confidence in the MAP label assignment, the labels for its corresponding variables at the fine grid should be inferred by minimizing the original energy. Our method computes uncertainty estimates associated with the label assignments (see figure 1e) and uses them to choose which regions (see figure 1f) of the image should be included in the optimization at the finer level. Experimental results show that this technique enables us to identify thin structures of the object which had been misclassified in the solution of the coarse energy. See figure 1 for an illustration.

Outline of the Paper In section 2 we define our notation and discuss the pairwise CRF formulation of the image segmentation problem. We explain the modus operandi of multi-scale energy minimization algorithms in section 3. In

section 4 we provide a detailed explanation of our methods for constructing the low resolution energy. This is followed by the explanation of our uncertainty driven strategy for isolating pixels which need to be solved at the higher resolution in section 5. The experimental results of our method are provided in section 6. We conclude by discussing related problems and future work in section 7.

2. Notation and Preliminaries

We start by defining the Markov and Conditional Random Field (MRF and CRF) models that are widely used for formulating labelling problems in vision. Consider a discrete random field $\mathbf{X}$ defined over a lattice $\mathcal{V} = \{1, 2, \ldots, n\}$ with a neighbourhood system $\mathcal{E}$. Each random variable $X_i \in \mathbf{X}$ is associated with a lattice point $i \in \mathcal{V}$ and takes a value from the label set $\mathcal{L} = \{l_1, l_2, \ldots, l_k\}$. The neighborhood system $\mathcal{E}$ of the random field is a set of edges (i, j) which connect some random variables. Any possible assignment of labels to the random variables will be called a *labelling* (denoted by $\mathbf{x}$) which takes values from the set $\mathbf{L} = \mathcal{L}^n$. The most probable or maximum a posterior (MAP) labelling $\mathbf{x}^*$ of the random field can be computed by maximizing the posterior or alternatively minimizing the energy of the random field as:

$$\mathbf{x}^{\text{opt}} = \arg\max_{\mathbf{x} \in \mathbf{L}} \Pr(\mathbf{x}|\mathbf{D}) = \arg\min_{\mathbf{x} \in \mathbf{L}} E(\mathbf{x}) \qquad (1)$$

2.1. Pairwise CRFs for Image Segmentation

The random field models used for most vision problems are *pairwise* i.e. their energy function $E : \mathcal{L}^n \to \mathbb{R}$ can be written as a sum of unary and pairwise functions as:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \phi_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \phi_{ij}(x_i, x_j). \qquad (2)$$

For image segmentation, the set $\mathcal{V}$ corresponds to the set of all image pixels, $\mathcal{E}$ is set of all edges between pixels in a 4 or 8 neighbourhood. The random variable X_i denotes the labelling of pixel i of the image. The label set $\mathcal{L}$ consists of two labels: foreground (fg) and background (bg). Every possible assignment of the random variables $\mathbf{x}$ (or configuration of the CRF) defines a segmentation.

The unary potential ϕ_i of the CRF is defined as the negative log of the likelihood of a label being assigned to pixel i. It can be computed from the colour of the pixel and the appearance models for the foreground and background segments[2]. Formally,

$$\phi(x_i = \mathcal{S}) = -\log \Pr(I_i | x_i = \mathcal{S}) \qquad (3)$$

---

[2]The appearance models are constructed from the user marked brush strokes as explained in [1, 23].

where $\mathcal{S} \in \{\text{fg,bg}\}$. The pairwise terms ϕ_{ij} of the CRF take the form of a contrast sensitive Ising model:

$$\phi(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j, \\ g(i, j) & \text{otherwise,} \end{cases} \qquad (4)$$

where the function $g(i, j)$ is an edge feature based on the difference in colors of neighboring pixels [3]. It is typically defined as:

$$g(i, j) = \theta_p + \theta_v \exp(-\theta_\beta ||I_i - I_j||^2), \qquad (5)$$

where I_i and I_j are the colour vectors of pixel i and j respectively. (see [3, 23] for more details).

2.2. MAP Inference in Pairwise CRFs

There are a number of standard algorithms for minimizing pairwise energy functions of the form (2). These can be classified into two broad categories: (a) message passing algorithms such as belief propagation and its tree reweighted versions (BP/TRW) [14, 28] and (b) combinatorial algorithms such as graph cuts [5, 15]. These algorithms allow for the exact or approximate minimization of pairwise energy functions in polynomial time. Some families of pairwise energy functions such as *submodular* functions can be minimized exactly in polynomial time. In fact they can be transformed to an st minimum cut problem (st-mincut) [2], which can be solved exactly in polynomial time using fast maximum flow (max-flow) algorithms. The energy function (2) for the two label image segmentation with Ising model pairwise potentials is submodular and hence can be exactly minimized in polynomial time.

3. Multi-scale Energy Minimization

We now provide an overview of multi-scale methods for energy minimization. These algorithms have the following basic steps:

Construction of a smaller problem A new energy function $E^l : \mathcal{L}^{n\omega_s} \to \mathbb{R}$ is constructed over a smaller grid $(\mathcal{V}^l, \mathcal{E}^l)$ where $\mathcal{V}^l$ denotes the set of lattice points, and $\mathcal{E}^l$ denotes the corresponding edge set. This grid has $|\mathcal{V}^l| = n\omega_s$ variables[3], where ω_s is the scaling parameter ($0 \le \omega_s \le 1$). Let $\mathbf{X}^l = \{X_i^l, i \in \mathcal{V}^l\}$ denote the vector of variables defined on $\mathcal{V}^l$. We will denote their labelling by $\mathbf{x}^l = \{x_1^l, x_2^l, ..., x_{n\omega_s}^l\}$.

Computation of a partial labelling The coarse energy E^l is minimized to extract a partial labelling $\mathbf{x}^*$ for the original random variables $\mathbf{X}$. Formally, each variables X_i is assigned one label from the extended label set $\mathcal{L} \cup \{\epsilon\}$. The assignment $x_i^* = \epsilon$ indicates that variable X_i has not been assigned any label.

---

[3]Recall that the original energy E had n variables.

Solving the Partial Labelling induced Projection The final solution of the original problem is obtained by minimizing a projection of the original energy function E. This energy projection $E' : \mathcal{L}^{n_\epsilon} \to \mathbb{R}$ is constructed from $E(\cdot)$ by fixing the values of the labelled variables as: $E'(\mathbf{x}) = E(\mathbf{x}_p)$ where n_ϵ is the number of unlabelled variables i.e. those assigned label ϵ.

3.1. Partial Labelling Quality

We will now discuss the question: *what is a good partial labelling?* If all variables in the partial solution $\mathbf{x}^*$ are labelled, then the projection E' of the energy will take no variables as argument [4] and would be trivially minimized. On the other hand, if all variables are unlabelled, the projection of the energy will be the same as the original energy ($E' = E$) and we would not obtain any speed-up. While constructing the partial labelling $\mathbf{x}^*$, we also want to make sure that all labelled variable are assigned the MAP label i.e. $\mathbf{x}_i^* \neq \epsilon \Rightarrow \mathbf{x}_i^* = \mathbf{x}_i^{\text{opt}}$. This will ensure that $\min E'(\mathbf{x}) = \min E(\mathbf{x})$.

We will measure the quality of a partial labelling using two measures: (1) Percentage of unlabelled variables (P_u) (lower the better), and (2) Percentage of correct label assignments (P_c). Formally, these are defined as:

$$P_u = \frac{100}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} (x_i^p \neq \epsilon), \text{ and } P_c = \frac{100 \sum_{i \in \mathcal{V}} (x_i^p = x_i^{\text{opt}} n)}{\sum_{i \in \mathcal{V}} (x_i^p \neq \epsilon)} \tag{6}$$

where $(z_1 \neq z_2)$ and $(z_1 = z_2)$ represent indicator functions which return 1 when their argument is true.

Computation Complexity Let us denote the complexity of the algorithm we are using to minimize the original energy E by $O(\mathcal{T}(n))$, where $\mathcal{T}(n)$ is any function of n. For instance, the complexity of max-flow based algorithms for minimizing submodular functions of the form (2) is $O(n^3)$, so $\mathcal{T}(n) = n^3$. The computation time for the multi-scale algorithm can be divided into two parts. **(1)** The time taken for computing the partial solution by minimizing the coarse energy E^l. More specifically, $\mathcal{T}(n\omega)$ for minimizing the energy over $n\omega$ variables and a linear term (n) for extracting the partial solution, thus resulting in the complexity $O(\mathcal{T}(n\omega) + n)$. **(2)** Time taken for minimizing the projection of the energy function which is $O(\mathcal{T}(P_u))$. This the final complexity is: $O(\mathcal{T}(n\omega) + n + \mathcal{T}(P_u))$.

In the next sections, we will discuss our strategy for constructing the smaller minimization problem and extracting a partial labelling $\mathbf{x}^*$ for the variables of the original energy function from it. For purposes of explanation, we will use the two label (fg/bg) image segmentation problem as an example. However, as stated earlier, our framework is general

---

[4]It will be a constant function

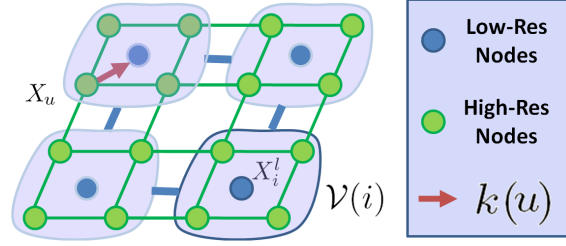

Figure 2. *The figure shows the mapping between the coarse and fine grids. The $4 \times 4 = 16$ variable problem is represented by a coarse 2×2 variable problem. The scaling parameter $\omega_s = 0.25$, while $\theta_s = 2$.*

and can be used to minimize energy functions corresponding to any labelling problem.

4. Constructing the Low Resolution Problem

We now explain how a smaller energy minimization problem over the coarse grid $\mathcal{V}^l$ is constructed from the original large scale problem which was defined over $\mathcal{V}$.

There is a many-one mapping between points in $\mathcal{V}$ and $\mathcal{V}^l$. We denote the set of indices of nodes in $\mathcal{V}$ which map to the node $i \in \mathcal{V}^l$ by $\mathcal{V}(i)$ which we will call the child set of i. We also define the function $k : \mathcal{V} \to \mathcal{V}^l$ which given a node i in the original grid, returns the index of its parent node in the reduced grid $\mathcal{V}^l$. For image labelling problems, the traditional approach is to map a square $\theta_s \times \theta_s$ grid of nodes in $\mathcal{V}$ to a single node in the small scale grid $\mathcal{V}^l$, where $\theta_s^2 = \frac{1}{\omega_s}$. This mapping is illustrated in figure 2. More recently, super-pixel based mappings have become increasingly popular which has led to work on the extraction of super-pixel lattices [20]. We will use a square patch mapping in our experiments. However, it should be noted that our framework allows the use of any many-one mapping between the fine and coarse problems.

The energy E^l defined over $\mathcal{V}^l$ has the same form as the original energy E (2). Formally, it can be written as:

$$E^l(\mathbf{x}^l) = \sum_{i \in \mathcal{V}^l} \phi_i^l(x_i^l) + \sum_{(i,j) \in \mathcal{E}^l} \phi_{ij}^l(x_i^l, x_j^l). \tag{7}$$

We now need to define the unary and pairwise potentials for this energy function.

4.1. Traditional Method

Traditional band-based multi-scale methods for image segmentation (such as those proposed in [19, 26]) define the energy potentials using a low-resolution version $\mathcal{I}^l$ of the original image $\mathcal{I}$. Formally, the potential functions are defined as: $\phi^l(x_i^l = \mathcal{S}) = -\log \Pr(I_i^l | x_i^l = \mathcal{S})$, where $\mathcal{S} \in \{\text{fg,bg}\}$. The pairwise potential is defined as:

$$\phi_{ij}^l(x_i^l, x_j^l) = \begin{cases} 0 & \text{if } x_i^l = x_j^l, \\ g^l(i, j) & \text{otherwise.} \end{cases} \tag{8}$$
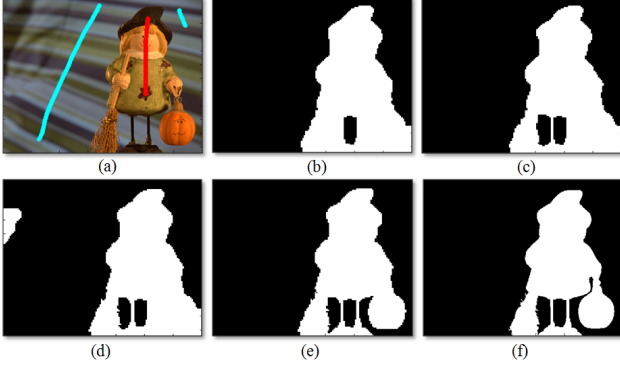
Figure 3. *Constructing the coarse energy. The figure shows the results of using different methods for constructing the coarse energy function.* **(a)** *The original image with user marked brush strokes for the different segments: foreground (red) and background(blue). The coarse level grid for this experiment is constructed with scaling parameter $\omega_s = 0.04$ ($\theta_s = 5$).* **(b)** *The solution obtained by constructing the energy using a low resolution version of the original image (as explained in section 4.1).* **(c)** *Solution of the energy with the scale-corrected parameter values for the pairwise potentials (computed as explained in section 4.2).* **(d)** *Segmentation obtained by using the energy constructed from the original energy (as explained in section 4.3).* **(e)** *Segmentation obtained by using the pairwise potential definition 12.* **(f)** *The solution obtained by minimizing the energy function (2) defined over the full-resolution image grid.*

Here the function $g^l(\cdot)$ is defined as:

$$g^l(i,j) = \theta_p^l + \theta_v^l \exp(-\theta_\beta^l ||I_i^l - I_j^l||^2), \qquad (9)$$

where I_i^l and I_j^l are the colour vectors of pixels of the low resolution image $\mathcal{I}^l$ which correspond to variables X_i^l and X_i^l respectively. The coarse segmentation result obtained by using the above potential function definitions is shown in figure 3b. The quantitative results of this method will be annotated with the symbol (**I**) (which indicates that the coarse scale energy was constructed from the low resolution image).

4.2. Scale Dependent Parameter Selection

We now address the problem of setting the parameters $\{\theta_p^l, \theta_v^l, \theta_\beta^l\}$ of the pairwise potentials of the coarse level energy E^l. Most band-based multi-scale methods overlook this issue and assume that the parameter values are the same as the ones used for defining the potentials 5 of the original energy E (2) i.e. $\{\theta_p^l, \theta_v^l, \theta_\beta^l\} = \{\theta_p, \theta_v, \theta_\beta\}$. Our experiments show that these parameter values result in over smooth segmentations for the low-resolution image.

An insight on the cause of this erroneous behaviour can be gained through the work of Boykov and Kolmogorov [4]. The values of the above mentioned parameters controls the strength of the pairwise potentials. These potentials in turn

penalize segmentation boundaries. The length of the segmentation boundary in pixel terms is reduced when we move from the original image $\mathcal{I}$ to the low-resolution image $\mathcal{I}^l$. This reduction is inversely proportional to $\theta_s = \frac{1}{\sqrt{\omega_s}}$. Thus, we would need to reduce the strength of the pairwise potentials by the same amount. In order to achieve this, we set the potential parameters as: $\{\theta_p^l, \theta_v^l, \theta_\beta^l\} = \{\sqrt{\omega_s}\theta_p, \sqrt{\omega_s}\theta_v, \theta_\beta\}$[5]. The segmentation result obtained by minimizing the scale corrected energy is shown in figure 3c. Quantitative results obtained by using these potential parameters will be marked with the symbol ($\mathbf{I}^c$).

4.3. Construction from the Original Energy

Unlike the multi-scale methods mentioned above, we define the potentials for the coarse grid energy E^l directly from the potentials of the original energy E (2). A simple method to compute the unary potential for a variable X_i^l is to sum the unary potentials of all the variables in its child set $\mathcal{V}(i)$. Formally, $\phi_i^l(x_i^l = \mathcal{S}) = \sum_{j \in \mathcal{V}(i)} \phi(x_j = \mathcal{S})$. Similarly, the pairwise potential for an edge $(u, v) \in \mathcal{E}^l$ is computed by summing the pairwise potentials defined between their children. Formally,

$$\phi_{ij}^l(x_i^l, x_j^l) = \begin{cases} 0 & \text{if } x_i^l = x_j^l, \\ g_e^l(i,j) & \text{otherwise.} \end{cases} \qquad (10)$$

Here the function $g_e^l(\cdot)$ is defined as:

$$g_e^l(i,j) = \sum_{\substack{(u,v) \in \mathcal{E}: u \in \mathcal{V}(i) \\ v \in \mathcal{V}(j)}} \theta_p + \theta_v \exp(-\theta_\beta ||I_u - I_v||^2). \quad (11)$$

The segmentation result obtained by minimizing the this coarse scale energy is shown in figure 3(d).

4.4. Incorporating the Internal Pairwise Potentials

At first glance, the above definition may seem to be a reasonable strategy. However, on careful analysis, we observe that it ignores the definitions of the pairwise potentials defined on variables X_i and X_j ($i, j \in \mathcal{V}$) which have the same parent i.e. $k(i) = k(j)$. In fact, it can be verified that this approximation is correct only if we assume that Ising model pairwise potentials with infinite cost are defined over every pair of variables X_i and X_j ($i, j \in \mathcal{V}$) which share the same parent.

This situation can be more easily understood by considering the maxflow problem corresponding to the original energy minimization problem. As an example, consider a multi-scale decomposition where variables in a 2×2 square on the original grid share the same parent (see figure 2). The pairwise potential definition (10) would translate into a

---

[5]Note that the parameter values depend on the topology of the graph, and this equation would be different for the 3D voxel segmentation problem.

capacity in the max-flow graph that would allow flow coming into any child node to pass through to any other child node and flow out from it. Obviously, this is a very bad assumption since in reality the child nodes in the graph corresponding to the original energy may be disconnected from each other[6]. This added phantom capacity would make the pairwise potentials very strong and result in over-smooth segmentations (as seen in figure 3d).

We resolve the problem of excess flow capacity by computing a lower bound on the flow that can be passed between child nodes constituting any two sides of the child-set square of a coarse variable $X_i^l (i \in \mathcal{V}^l)$. This capacity is used as the upper bound on the capacity of the edges which connect a particular parent node to other parent nodes. We estimate the lower bound by finding the minimum capacity edge in the child set.

Coming back to the energy formulation, our pairwise potential has the same form as (10) but now we define function

$$g_e^l(i,j) = \mathcal{R}(i,j) \min_{\substack{k \in \{i,j\}, (u,v) \in \mathcal{E} \\ u \in \mathcal{V}(k), v \in \mathcal{V}}} \theta_p + \theta_v \exp(-\theta_\beta ||I_u - I_v||^2).$$
(12)

where $\mathcal{R}(i,j)$ is the number of edges between child-sets of the two coarse level variables X_i and X_j $(i,j \in \mathcal{V}^l)$. Formally,

$$\mathcal{R}(i,j) = \sum_{(u,v) \in \mathcal{E} : u \in \mathcal{V}(i), v \in \mathcal{V}(j)} 1.$$
(13)

The segmentation result obtained by minimizing the this coarse scale energy is shown in figure 3e. Quantitative results obtained by using these potential function definition will be marked by the symbol ($\mathbf{E}^c$).

5. Computing Partial Labellings

We now explain our method for computing a partial labelling from the coarse energy E^l constructed in the previous section.

5.1. Partial Labellings from Boundary Bands

Conventional multi-scale methods use the lowest cost solution $\mathbf{x}^{l*}$ of the coarse energy E^l for defining the partial labelling. For instance, the method proposed in [19] first defined a full labelling of the original variables $\mathbf{x}^e$ as: $x_i^e = x_{k(i)}^{l*}$, where recall $k(i)$ returns the parent of any variable $X_i, i \in \mathcal{V}$.

From the solution $\mathbf{x}_e$, the method then computes a boundary distance function $\mathcal{D} : \mathcal{V} \to \mathbb{R}$ which returns the distance of a pixel i from the boundary between the different label segments (fg and bg) in the labelling $\mathbf{x}_e$. The

---

[6]This is true if the Ising model penalty for taking different labels is zero.

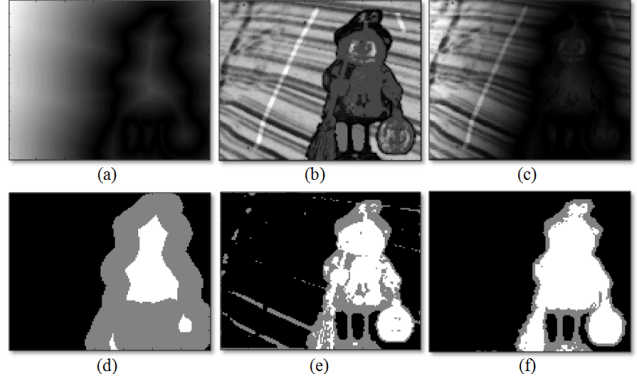Figure 4. *Computing the partial labelling. The figure shows the results of using different techniques for computing the partial labelling. For this experiment, we used the image shown in figure (3a). We constructed an energy on a coarse level grid with scaling parameter $\omega_s = 0.04$ ($\theta_s = 5$) using the method explained in section 4.3. The result of the minimization has been shown in figure (3e). Images (a), (b) and (c) depict the response of the boundary distance $\mathcal{D}(\cdot)$, confidence score $\mu(\cdot)$, and hybrid score $\mathcal{H}(\cdot)$ respectively. These functions were used to obtain the partial solutions shown in images (d), (e) and (f) is the same order. The thresholds for marking pixels as unlabelled was choosen to ensure all marked pixels took the MAP label of the original energy. It can be seen that the hybrid approach requires less number of unlabelled pixels compared to the confidence function which in turn requires less pixels compared to the boundary band approach.*

set of variables assigned the label ϵ is computed by finding the set of nodes $\mathcal{P}_B(\delta)$ which are at most a particular distance δ away from the boundary i.e. i.e. $x_i^e = \epsilon, \forall i \in \mathcal{V}, \mathcal{D}(i) \leq \delta$. Formally, the set is defined as $\mathcal{P}_B(\delta) = \{i : i \in \mathcal{V}, \mathcal{D}(i) \leq \delta\}$. The quality of partial labellings obtained from boundary bands of the coarse segmentation results shown in figure 3e can be seen in the graph shown in figure 5.

This band-based approach for extracting partial labellings does not take into account the confidence or uncertainty associated with the label assignment for any variable $x_i^l, i \in \mathcal{V}^l$. Next, we show how you can compute partial solutions using min-marginals as a confidence measure.

5.2. Partial Labelling from Min-marginals

Given an energy function, the min-marginal encodes the confidence associated with a variable being assigned the MAP label. More concretely, the min-marginal $\psi_{i;a}$ returns the energy value obtained by fixing the value of variable X_i to label a ($x_i = a$) and minimizing over all remaining variables. Formally, $\psi_{i;a} = \min_{\mathbf{x}, x_i = a} E(\mathbf{x})$.

Min-marginals as an Uncertainty Measure Min-marginals naturally encode the uncertainty of a labelling and have been successfully used for solving a number of vi-

sion and learning problems [8, 9]. The exact min-marginals associated with graph cut solutions can be efficiently computed using dynamic graph cuts in roughly 3-4 times the time taken for minimizing the energy [12, 13]. This method can handle all graph representable submodular energy functions [2, 11, 24] which are widely used for image and 3D volume segmentation problems.

In their recent work on computing optical flow Glocker *et al.* [9] proposed a method which can be used to compute approximate min-marginals of non-submodular energy functions. Such approximate estimates can also be computed using variants of belief propagation and tree reweighted message passing.

Isolating Pixels with High Uncertainty   We use the absolute difference between the min-marginals corresponding to the fg and bg labels as our confidence score function $\mu : \mathcal{V} \to \mathbb{R}$. Formally, $\mu(i) = ||\psi_{i;fg} - \psi_{i;bg}||$. If the difference between min-marginals of any variable X_i corresponding to taking the MAP label and any other label is large, then the variable is assigned a high confidence score. The set of variables assigned the label ϵ in the partial solution is now computed by finding the set of nodes $\mathcal{P}_M(\delta)$ whose confidence scores are less than some constant δ_μ i.e. $x_i^e = \epsilon, \forall i \in \mathcal{V}, \mu(i) \leq \delta_\mu$. Formally, the set is defined as $\mathcal{P}_M(\delta_\mu) = \{ i : i \in \mathcal{V}, \mu(i) \leq \delta_\mu \}$. Similar to the boundary band-width parameter δ_B, the value of the confidence threshold δ_μ can be used to the change the number of unlabelled variables. (see figure 4e)

Hybrid Selection Measures   Although, the min-marginals based confidence function is able to obtain good partial labellings, we observed that it sometimes selects variables which are spatially distant from the main foreground segment. This motivated us to test a new hybrid measure which combines the boundary and uncertainty based techniques described above. We construct the new function $\mathcal{H} : \mathcal{V} \to \mathbb{R}$ which is defined as: $\mathcal{H}(i) = \mu(i)\mathcal{D}(i)$. As before, the set of variables assigned the label ϵ is now computed by finding the set of nodes $\mathcal{P}_H(\delta) = \{ i : i \in \mathcal{V}, \mu(i) \leq \delta_\mu \}$. Formally, the partial solution is defined as: $x_i^e = \epsilon, \forall i \in \mathcal{V}, \mathcal{H}(i) \leq \delta_H$. (see figure 4f)

6. Experiments

We will now describe the qualitative and quantitative results of our experiments. We test our method on the image segmentation problem using the data set used in [22].

Relating Speed with Accuracy   The speed and accuracy of a multi-scale method are inversely proportional to each other. The correctness of the partial labellings can be easily changed by changing the threshold parameters δ_B, δ_μ,
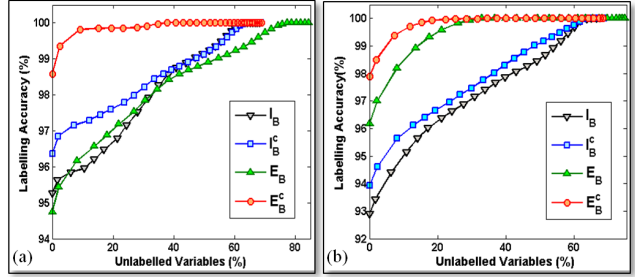

Figure 5. *Results of different multi-scale energy constructions. The graphs show how the accuracy P_c of a partial solution changes as we increase the percentage of unlabelled variables (P_u). Graphs (a) and (b) show the results of using the band based approach on the solutions generated from different energy construction methods.* **Key: $\mathbf{I_B}$** : *result of energy constructed from the low resolution image (see section 4.1),* $\mathbf{I_B^c}$: *same energy with scale dependent parameters (section 4.2),* $\mathbf{E_B}$: *smaller problem constructed from the original energy (section 4.3),* $\mathbf{E_B^c}$: *smaller problem constructed using the lower bound on pairwise potentials (section 4.4).*

and δ_H. For instance, setting $\delta_B = \sqrt{I_{width}^2 + I_{height}^2}$ will make sure that all variables in the partial solution are unlabelled. Here I_{width} and I_{height} are the width and height of the image I to be segmented respectively. However, this would also nullify the benefit of the multi-scale technique as we will need to minimize the energy over all the variables.

The key matter we want to investigate is how the percentage of variables (P_u) unlabelled in the partial solutions produced by the different multi-scale minimization techniques affect correctness P_c of the solution. We divide our experiments into two parts to investigate how the performance is affected by the use of different: **(1)** Methods for constructing the smaller energy minimization problem (section 4), **(2)** Methods for extracting the partial labelling from the smaller energy (section 5).

Comparing Energy Construction Methods   We compared the quality of partial labellings generated from different coarse energy constructions using the boundary band method. The results for the images shown in figure 1(a) and 3(a) are shown in graphs in figure 5(a) and (b). It can be seen from the results that using scale dependent parameters is better than the traditional approach. Further, the method for constructing coarse energy directly from the original energy function outperforms other methods. It is able to achieve a correctness of $P_c = 99.5\%$ with less than 10% of unlabelled variables.

Comparing Methods for Partial Solution Extraction
The relative performance of different techniques for extracting the partial solution is now analyzed. Consider the problem of segmenting the image shown in figure 3(a). Figure
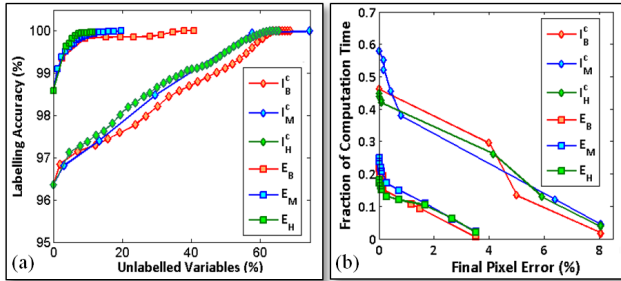
Figure 6. *Computation time and accuracy of different multi-scale methods. Graph (a) shows how the accuracies P_c of partial solutions extracted using different methods change as we increase the percentage of unlabelled variables (P_u). It depicts the results of using different partial solution extraction methods. They key is the same as the one used in graph 5. Subscripts B, M and H denote that the partial solutions were extracted using boundary distance, min-marginal based uncertainty, and the hybrid uncertainty boundary bands respectively. Graph (b) shows the fraction of computation required to achieve a particular pixel labelling accuracy in the final segmentation solution.*

4 shows the different partial labellings extracted from the coarse energy defined in section 4.4. The size of the sets $\mathcal{P}_B$, $\mathcal{P}_M$, and $\mathcal{P}_H$ was choosen to ensure that the partial labelling were fully correct ($P_c = 100\%$). This ensured that they led to the exact optimal solution of the original problem. The percentage of unlabelled variables required for the boundary band, uncertainty, and hybrid approaches were 35.29%, 17.36%, and 9.03% respectively. The results on the image shown in figure 1(a) are shown in figure 6(a). For dearth of space we only provide results with the methods described in section 4.2 and 4.4 because they performed better in all experiments. It can be seen that the hybrid partial labelling technique results in a much smaller problem to be solved while still obtaining the exact global minimum.

Relating Computational Speedup and Accuracy We now discuss the speed-up obtained by our multi-scale methods. As explained in section 3, the total computation time $\mathcal{T}$ of a multi-scale method has two primary components: time for partial solution computation t_s, and that for solving the resulting projection (t_p). The size of the projection (and thus t_p) is dependent on the level of accuracy required by the user, while t_s is independent.

For the boundary band method, t_s is equal to the time needed to minimize the coarse energy. For the min-marginal based confidence and hybrid extraction methods, t_s is the time needed to find the min-marginals, which is a much more expensive operation. For instance, for the image shown in figure 1(a), it takes only 1 msec to minimize the coarse energy (from 4.2), while it takes 10 msec to compute all the min-marginals. However, for any given solution accuracy, the min-marginal based methods produce a smaller

partial solution compared to the boundary band method. For high levels of accuracy, the size of the projection is large and thus t_p is the dominant time. Thus, min-marginals based methods are able to out-perform band based methods. However, for low levels of accuracy, the size of the projection is very small, which makes t_s to dominate. In such cases, the boundary band based approach outperforms the min-marginals based approach. The performance of all the methods can be seen in the graph shown in figure 6(b).

7. Discussion and Conclusions

In this paper, we presented a uncertainty driven approach for multi-scale energy minimization. We showed that this strategy allows us to compute solutions close to the globally optimal in a fraction of the time required by a conventional energy minimization algorithm. It allowed us to solve large scale instances of the two label interactive image segmentation problem extremely quickly. The method proposed in this paper is general and can be applied to any labelling problem. In future work we would like to investigate the how general energies defined over variables with large label sets can be minimized in a multi-scale fashion. In this regard the work of [9] provides a nice initial direction.

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. 3

[2] E. Boros and P. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 2002. 3, 7

[3] Y. Boykov and M. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *ICCV*, pages I: 105–112, 2001. 1, 3

[4] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *ICCV*, pages 26–33, 2003. 5

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 2001. 1, 3

[6] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 1983. 2

[7] A. Delong and Y. Boykov. A scalable graph-cut algorithm for n-d grids. In *CVPR*, 2008. 1

[8] J. Duchi, D. Tarlow, G. Elidan, and D. Koller. Using combinatorial optimization within max-product belief propagation. In *NIPS*, pages 369–376, 2006. 7

[9] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab. Optical flow estimation with uncertainties through dynamic mrfs. In *CVPR*, 2008. 7, 8

[10] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *RoyalStat*, B: 51(2):271–279, 1989. 1

[11] H. Ishikawa. Exact optimization for markov random fields with convex priors. *PAMI*, 25:1333–1336, October 2003. 1, 7

[12] P. Kohli and P. Torr. Efficiently solving dynamic markov random fields using graph cuts. In *ICCV*, volume II, pages 922–929, 2005. 7

[13] P. Kohli and P. Torr. Measuring uncertainty in graph cut solutions: Efficiently computing min-marginal energies using dynamic graph cuts. In *ECCV*, pages 30–43, 2006. 7

[14] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1568–1583, 2006. 3

[15] N. Komodakis, G. Tziritas, and N. Paragios. Fast, approximately optimal solutions for single and dynamic MRFs. In *CVPR*, 2007. 3

[16] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Trans. Graph.*, 2007. 1

[17] J. Kopf, M. Uyttendaele, O. Deussen, and M. F. Cohen. Capturing and viewing gigapixel images. *ACM Trans. Graph.*, 2007. 1

[18] V. S. Lempitsky and Y. Boykov. Global optimization for shape fitting. In *CVPR*, 2007. 1, 2

[19] H. Lombaert, Y. Sun, L. Grady, and C. Xu. A multilevel banded graph cuts method for fast image segmentation. In *ICCV*, 2005. 1, 2, 4, 6

[20] A. P. Moore, S. Prince, J. Warrell, U. Mohammed, and G. Jones. Superpixel lattices. In *CVPR*, 2008. 4

[21] P. Pérez and F. Heitz. Restriction of a markov random field on a graph and multiresolution statistical image modeling. *IEEE Transactions on Information Theory*, 1996. 1

[22] C. Rhemann, C. Rother, A. Rav-Acha, and T. Sharp. High resolution matting via interactive trimap segmentation. In *CVPR*, 2008. 7

[23] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: interactive foreground extraction using iterated graph cuts. In *SIGGRAPH*, pages 309–314, 2004. 3

[24] D. Schlesinger and B. Flach. Transforming an arbitrary minsum problem into a binary one. Technical Report TUD-FI06-01, Dresden University of Technology, April 2006. 7

[25] S. N. Sinha, P. Mordohai, and M. Pollefeys. Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh. In *ICCV*, 2007. 1

[26] A. K. Sinop and L. Grady. Accurate banded graph cut segmentation of thin structures using laplacian pyramids. In *MICCAI*, 2006. 2, 4

[27] G. Vogiatzis, P. Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts. In *CVPR*, 2005. 1

[28] Y. Weiss and W. Freeman. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *Transactions on Information Theory*, 2001. 3