A Multiscale Hybrid Model Exploiting Heterogeneous Contextual Relationships for Image Segmentation

Lei Zhang and Qiang Ji
Rensselaer Polytechnic Institute
110 8th St., Troy, NY 12180
zhangl2@rpi.edu, qji@ecse.rpi.edu

Abstract

We propose a framework that can conveniently capture heterogeneous relationships among multiple random variables. The framework is formulated based on a hybrid probabilistic graphical model. It allows using both directed links and undirected links to capture various types of relationships. Based on this framework, we develop a multiscale hybrid model for image segmentation. The multiscale model systematically captures the spatial relationships and causal relationships among such image entities as regions, edges, and vertices at different scales. We further show how to parameterize such a hybrid model and how to factorize its joint probability distribution according to the global Markov properties. Based on this factorization, we exploit the Factor Graph theory to perform joint probabilistic inference and solve for the image segmentation problem.

1. Introduction

Image segmentation is an important low level vision problem. It provides the basis for other middle level or high level problems such as object recognition, scene understanding, etc. In image segmentation, we deal with different image entities such as pixels, regions, edges, junctions, etc. Researchers have noticed that it is very important to exploit the relationships between image entities for solving the image segmentation problem. For example, it is hard to distinguish the regions in the squares of Figure 1 purely depending on their individual appearances. However, the relationships of these regions to other entities such as their nearby neighbors or long-range neighbors can help to disambiguate the problem.

Previously, researchers have incorporated different relationships as additional knowledge besides the image data. They have included the global shape constraints, spatial relationships, smoothness constraints, scene contexts, consistency of image labels at multi-scales [21, 18, 9, 2, 12, 7], etc. Incorporation of these information has been demonstrated to help solve the specific problem.

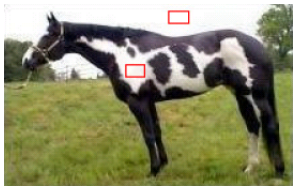


Figure 1. An example illustrating the importance of the contextual relationships. Without contextual information, it is hard to discriminate the regions within the rectangles.

The natural relationships between different image entities are often heterogeneous. Some relationships can be naturally interpreted as causal relationships. Other relationships like the correlations or mutual interactions do not have explicit causal meaning. All these heterogeneous relationships can be useful and informative. The problem is how to model them in a systematic way. We need a powerful tool to address this modeling problem.

In machine learning community, Probabilistic Graphical Models (PGMs) have been developed as a powerful modeling tool. The PGM is a marriage between the graph theory and statistics. It provides a systematic way to model various relationships and is intuitively easy to understand. In addition, principled methods have been developed to perform probabilistic inference in a PGM to search for the optimal states of random variables, given the evidence.

Traditionally, different types of PGMs are developed to model certain kinds of relationships. These PGMs can be divided into the undirected PGMs and the directed acyclic PGMs. The Markov Random Fields (MRF) [9, 2, 20] and Conditional Random Fields (CRF) [16, 15] mainly capture the non-causal relationships such as the spatial correlation. On the other hand, the Bayesian Network (BN) [24] [29] [7] and Hidden Markov Model (HMM) mainly model the causal relationships. Both of them have been exploited to solve computer vision problems.

Although the undirected PGMs and directed PGMs can individually capture certain types of relationships, they can-

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

not model heterogenous relationships. Since heterogenous contextual relationships extensively exist in real problems, there is a need for a single framework that can simultaneously capture all these relationships in a systematic way.

In this paper, we present such a framework that is capable of capturing heterogeneous relationships and perform inference in a principled way. To demonstrate the capability of this framework, we apply it to the figure/ground image segmentation problem. Our main contributions include: 1) we propose a multiscale hybrid probabilistic graphical model to represent multiple image entities and capture the heterogeneous contextual relationships among them at different scale levels; 2) we show how to factorize the joint probability distribution according to the graphical structure of the hybrid model and how to parameterize such a model; 3) based on the factorization, we demonstrate how to use the Factor Graph theory to perform joint probabilistic inference in a principled way to solve the problem.

2. Overview

Our model is built on a multiscale framework, as shown in Figure 2. There are different image entities in image segmentation, such as regions and edges. Traditionally, image segmentation can be treated as a problem to label each pixel of the image. The group of connected pixels with the same label form a segmented region. This type of method is classified into the region-based image segmentation. On the other hand, image segmentation can also be treated as a problem to find the boundary of the object of interest. This type of method is the edge-based image segmentation. Besides, there are other image entities that may also be useful. For example, edges can intersect to form vertices (junctions). In this paper, a vertex means the intersection of multiple (more than two) edge segments. These entities also provide useful information for image segmentation. The vertex implies the intersection of multiple regions.

There are many heterogeneous contextual relationships between image entities that naturally exist and are informative. We capture these useful relationships in the multiscale framework through different types of links. Figure 3 illustrates those captured contextual relationships.

There are some natural causalities between these image entities. First, two adjacent regions intersect to form an edge. If these two regions have different labels, they form/cause a boundary between them. Second, multiple edges intersect to form a vertex. Third, the region labels at the fine layer induce the the region labels at the coarse layer. In the mutliscale framework, the image labels that are consistent in different scales tend to be the reliable labels. Traditionally, the prior works assume the inter-layer relationships can be formulated as a coarse-to-fine Markov chain. Such a framework tries to propagate the image labels from the coarse layer to the fine layer. Different from these

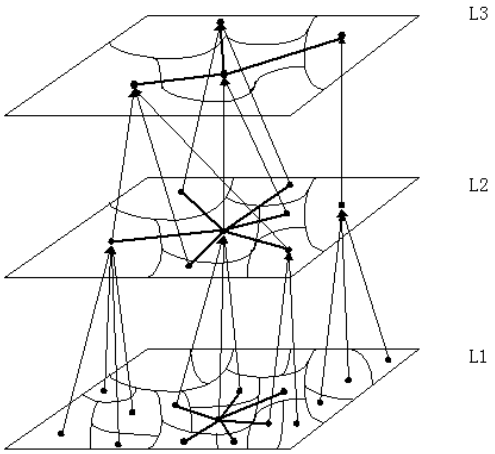


Figure 2. The multiscale framework consists of the inter-layer directed links pointing from the fine layer to the next coarse layer. The intra-layer undirected links (thick lines) represent the spatial correlations between region labels. For clarity, not all inter-layer and intra-layer links are drawn.

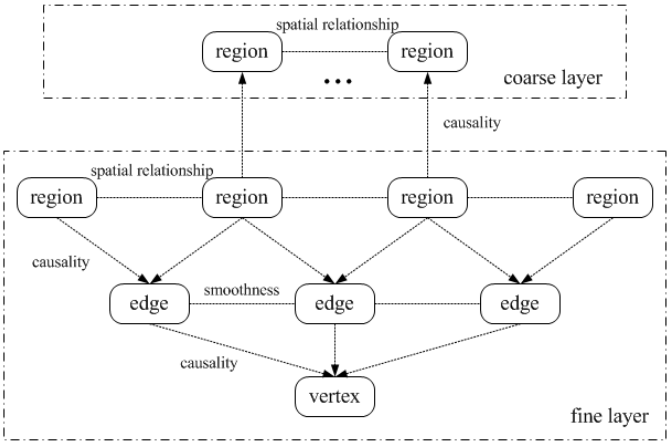


Figure 3. Image entities and the heterogeneous contextual relationships among them.

prior works, we think the image labels at different scales naturally form a kind of causal relationships since the image labels at the fine layer vote to predict the image labels at the next coarser layer.

Besides those causalities, there are some other useful contextual relationships. In image segmentation, the spatial correlations between image labels have often been exploited to encourage the local homogeneity of image labels. These spatial relationships are naturally modeled by undirected links. In addition, we observe that the boundary of a natural object is normally smooth. The connection between two connected edge segments therefore should be locally smooth. We enforce the smooth connection by the undirected links between the edge nodes. These links require the edge labels to be consistent with each other and form a smooth contour along the object boundary.

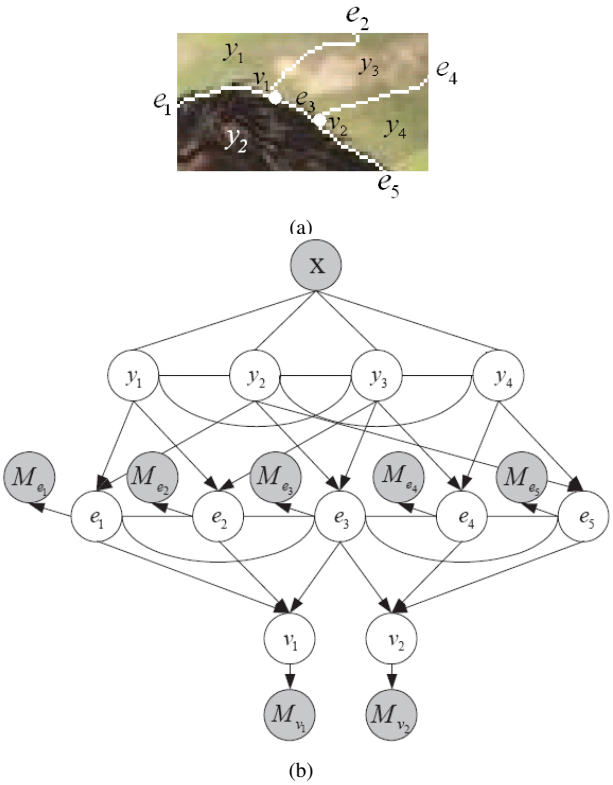


Figure 4. (a)A part of the oversegmentation. (b)A single-scale hybrid model capturing the heterogeneous contextual relationships.

3. A Hybrid Model for Image Segmentation

The proposed multiscale hybrid graphical model has a similar model structure at each scale, as shown in Figure 4. To construct the model at each scale, we first oversegment the image to produce a map of oversegmentation. Figure 4(a) shows a small part of this oversegmentation. From this map, we automatically find the small regions (hereafter referred to as superpixels) $\{y_i\}$, the edge segments $\{e_j\}$, and the vertices $\{v_k\}$. We then construct a hybrid model in Figure 4(b) to model these image entities and capture their heterogeneous relationships. In addition, each image entity has its own measurements that are represented by the shaded circles. The measurements of region nodes are represented by feature vectors $\mathbf{x}$ extracted from the observed image. The measurements of edge nodes can be defined similarly. In this paper, we simply use the average gradient magnitude as the edge measurement. The orientation of the edge segment is another useful information. It will be used to calculate the angle between two edge segments. This angle will be used to enforce the smoothness constraint. The measurement of the vertex is discretized according to the Harris corner response [10].

Specifically, the hybrid model consists of region nodes $\mathbf{y} = \{y_i, i = 1..n\}$, edge nodes $\mathbf{e} = \{e_j, j = 1..m\}$, vertex nodes $\mathbf{v} = \{v_t, t = 1..w\}$ and their measurements $\{x_i, i = 1..n\}$, $\mathbf{M_e} = \{M_{e_j}, j = 1..m\}$ and $\mathbf{M_v} = \{M_{v_t}, t = 1..w\}$. All nodes except the measurements x_i and M_{e_j} are discrete nodes. $y_i \in \{1, -1\}$ and the state 1 means this region is a foreground. e_j and v_t are binary nodes. $e_j = 1$ means this edge segment belongs to the object boundary, and vice versa. $v_t = 1$ means the vertex is actually a corner along the object boundary.

The relationships among these nodes are captured by either directed or undirected links. The directed links represent the causalities between image entities, while the undirected links represent their spatial relationships (or mutual interaction). Specifically, the directed links between two different image entities (e.g. between regions and edges), between the same image entity at different scales, and between an image entity and its measurement, model the causal relationships. The undirected links between the same type of image entities at the same scale model their local spatial relationships or the local smoothness constraint.

4. Factorization in the Hybrid Model

Given the hybrid graphical model, we need factorize the joint probability distribution (JPD) of the random variables according to the graphical structure. We first show how to factorize the JPD in a single-scale hybrid model. Next, we will show how to factorize the JPD in the whole multiscale hybrid model.

4.1. Factorization in the Single-scale Hybrid Model

Based on the graphical structure of the hybrid model in Figure 4, we can factorize the JPD of all the nodes according to the conditional independence relationships encoded in the graphical structure. Our factorization is based on the Global Markov Property (GMP) of a hybrid graphical model (cf. Chapter 3 of [17]).

Direct application of the GMP theory requires finding the moral graph and the ancestral set [17]. For a complex hybrid graphical model, it is not so straightforward to find the moral graph and the ancestral set. It is therefore difficult to directly apply the GMP for factorization. However, Buntine [4] presents a simpler way to find out the general formulation of factorizing the JPD represented by a hybrid graph G. This method is based on the concepts of component subgraphs and master graph. We briefly recall the definitions of these concepts:

- *Definition 1:* Given a chain graph $\mathcal{G}$ over some variables $\mathcal{X}$, the *component subgraphs* are a coarser partition of variables $\mathcal{X}$ than the chain components, and are the coarsest partition where the set of subgraphs induced by the partition are connected, undirected or directed (but not mixed) subgraphs of the chain graph.

- *Definition 2:* The *master graph* is a directed graph $\mathcal{G}_\mathcal{M}$ whose nodes are component subgraphs and arcs con-

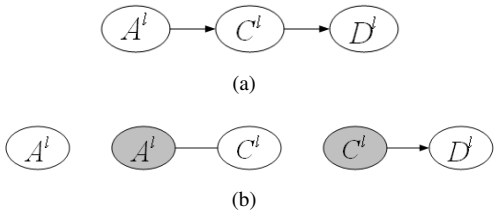


Figure 5. (a) the master graph of the single-scale hybrid model. (b) the component subgraphs of the single-scale hybrid model.

nect two component subgraphs $\mathcal{U}_i$ and $\mathcal{U}_j$ if a variable in $\mathcal{U}_i$ has a child in $\mathcal{U}_j$ in the graph $\mathcal{G}$.

This method basically works as follows: a hybrid model is first represented by its master graph and component subgraphs. Based on the master graph, the JPD of the hybrid model is decomposed into the product of functions defined on component subgraphs. Since each component subgraph is not a hybrid model, its probability function can be easily factorized into products of potentials or conditional probabilities. Buntine has proved the equivalence of this method to the Global Markov Property (cf. Theorem 2 in [4]). We use it to factorize the JPD of the proposed hybrid model.

For notational simplicity, we use the superscript (l) to indicate that the random variables $\mathbf{y}, \mathbf{x}, \mathbf{e}, \mathbf{v}, \mathbf{M_e}, \mathbf{M_v}$ belong to the lth layer. Let A^l denotes the variables in the undirected part of the lth layer, i.e. $A^l = \{\mathbf{y}^{(l)}, \mathbf{x}^{(l)}\}$, where $\mathbf{y}^{(l)}$ represents all the region nodes at the lth layer and $\mathbf{x}^{(l)}$ represent all region features calculated from the down-sampled image at this scale. Similarly, let B^l denotes all the random variables in the directed part of the lth layer, i.e. $B^l = \{\mathbf{e}^{(\mathbf{l})}, \mathbf{M}_\mathbf{e}^{(\mathbf{l})}, \mathbf{v}^{(\mathbf{l})}, \mathbf{M}_\mathbf{v}^{(\mathbf{l})}\}$. Let $C^l = \{\mathbf{e}^{(\mathbf{l})}\}$ and $D^l = \{\mathbf{M}_\mathbf{e}^{(\mathbf{l})}, \mathbf{v}^{(\mathbf{l})}, \mathbf{M}_\mathbf{v}^{(\mathbf{l})}\}$ denote the subset of nodes in $B^l = C^l \cup D^l$.

Figure 5(a) shows the master graph of the single-layer hybrid model. Its component subgraphs are shown in Figure 5(b). Based on the master graph and Butine's theory [4], the JPD is factorized as

$$P(A^l, B^l) = P(A^l, C^l, D^l) = P(D^l|C^l)P(C^l|A^l)P(A^l) \quad (1)$$

Each term in Eq.(1) can be formulated according to the component subgraphs. The factorization of $P(D^l|C^l)$ directly follows the d-separation rule of a directed acyclic graph [23], i.e.

$$\begin{aligned} P(D^l|C^l) &= P(\mathbf{M}_\mathbf{e}^{(\mathbf{l})}, \mathbf{v}^{(\mathbf{l})}, \mathbf{M}_\mathbf{v}^{(\mathbf{l})}|\mathbf{e}^{(\mathbf{l})}) \quad (2)\\ &= \prod_{e_j \in C^l} P(M_{e_j}|e_j) \prod_{v_t \in D^l} P(v_t|pa(v_t))P(M_{v_t}|v_t) \end{aligned}$$

where $pa()$ denotes the parents of a node.

According to the component subgraph that consists of A^l and C^l, we choose the conditional probability $P(C^l|A^l)$ as the following formulation,

$$P(C^l|A^l) \propto \prod_{e_j \in C^l} P(e_j|pa(e_j)) \prod_{<j,k>,(e_j,e_k)\in C^l} h(e_j, e_k) \quad (3)$$

where e_j and e_k correspond to the adjacent edge segments in the lth layer. The function $h(e_j, e_k)$ is a pairwise potential function that measures the compatibility of the edge labels according to the smoothness constraint. It depends on the angle ω_{jk} between the edges e_j and e_k. The function $h(e_j, e_k)$ is defined as

$$\begin{aligned} h(e_j, e_k) &= \{\alpha\delta(e_j)\delta(e_k) + (1-\alpha)[1-\delta(e_j)\delta(e_k)]\} \cdot \delta(\omega_{jk} < \omega^*)\\ &\quad + (1-\alpha)\delta(\omega_{jk} \geq \omega^*) \quad (4) \end{aligned}$$

where δ is the indicator function and $\omega^* = \frac{\pi}{6}$ is a threshold of the small angle. α is a weight to balance the penalty and set as 0.1. This definition penalizes the case that two edge segments along a small angle are both labeled as true. Otherwise, a sharp corner may exist in the boundary. It therefore encourages a smooth connection between edges.

With the factorization in Eq.(2) and Eq.(3), the conditional probability $P(B^l|A^l)$ can be finally factorized as

$$\begin{aligned} P(B^l|A^l) &= P(D^l|C^l)P(C^l|A^l) \quad (5)\\ &\propto \prod_{e_j \in C^l} P(M_{e_j}|e_j) \prod_{v_t \in D^l} P(v_t|pa(v_t))P(M_{v_t}|v_t)\\ &\quad \prod_{e_j \in C^l} P(e_j|pa(e_j)) \prod_{<j,k>,(e_j,e_k)\in C^l} h(e_j, e_k) \end{aligned}$$

In addition, we borrow the idea of Conditional Random Field to model the spatial correlations between region labels. We directly model the posteriori probability distribution of the region labels using potential functions, i.e.

$$P(y|x) \propto \prod_{i \in V} \phi(y_i, x_i) \prod_{i \in V} \prod_{j \in \mathcal{N}_i} exp(y_i y_j \lambda^T g_{ij}(x)) \quad (6)$$

where V is the set of all region nodes and y is the joint labeling of all region nodes. $\mathcal{N}_i$ denotes the neighborhood of the i^{th} region, which is automatically detected from the topological relationships among the regions. λ is the parameter vector. $g_{ij}(\cdot)$ represents the feature vector for a pair of region nodes i and j.

The first part $\phi(y_i, x_i)$ is the unary potential, which labels the i^{th} region according to the local features. It indicates how likely the i^{th} region will be assigned the label y_i given the local features x_i. We use a discriminative classifier based on a multi-layer perceptron (MLP) to define the unary potential, which is similar to [11].

The second part $exp(y_i y_j \lambda^T g_{ij}(x))$ is the pairwise potential that defines the interactions between region labels. We use a log-linear model to define this potential, which depends on the inner product of the weight vector λ and the pairwise feature vector $g_{ij}(x)$. The pairwise feature vector $g_{ij}(x)$ is defined as $[1, \ |x_i - x_j|]^T$, where $|\cdot|$ is the component-wise absolute value operator. Similar definitions have been used in [15].

Substituting Eq.(5) and Eq.(6) into Eq.(1), the JPD of all

nodes in the lth layer is factorized as

$$
\begin{aligned}
& P(A^l, B^l) \qquad (7)\\
&= P(B^l|A^l)P(\mathbf{y}^{(l)}|\mathbf{x}^{(l)})P(\mathbf{x}^{(l)})\\
&= \frac{1}{Z^l}\prod_{e_j\in C^l} P(M_{e_j}|e_j)\prod_{v_t\in D^l} P(v_t|pa(v_t))P(M_{v_t}|v_t)\\
&\quad \prod_{e_j\in C^l} P(e_j|pa(e_j))\prod_{<j,k>,(e_j,e_k)\in C^l} h(e_j,e_k)\\
&\quad \prod_{i\in V}\phi(y_i,x_i)\prod_{i\in V}\prod_{j\in\mathcal{N}_i}\exp(y_iy_j\lambda^T g_{ij}(\mathbf{x}))
\end{aligned}
$$

where Z^l is a normalization constant. Since the image $\mathbf{x}^{(l)}$ is observed, $P(\mathbf{x}^{(l)})$ becomes a constant and is therefore merged into the normalization constant.

Among these factorized terms, $P(M_{e_j}|e_j)$ and $P(M_{v_t}|v_t)$ are the likelihood models of the measurements of edges and vertices. $P(v_t|pa(v_t))$ and $P(e_j|pa(e_j))$ are the conditional probabilities of the vertex nodes and the edge nodes. The term $h(e_j, e_k)$ is the smoothness term defined in Eq.(4). The remaining terms $\phi(y_i, x_i)$ and $\exp(y_iy_j\lambda^T g_{ij}(\mathbf{x}))$ are the potential functions.

4.2. Factorization in the Multiscale Hybrid Model

The multiscale hybrid model consists of several directed parts and undirected parts. Its factorization of the JPD is more complex and requires more derivations. We can factorize the JPD based on the Global Markov Property. For convenience, we still use the method in [4] to factorize the joint probability. The master graph of the multiscale hybrid model (with 3 layers) is represented as Figure 6(a). Among these layers, the layer 1 corresponds to the finest layer. The component subgraphs of the multiscale hybrid model are shown in Figure 6(b).

According to the master graph, the joint probability in the multiscale hybrid model with 3 layers can be first coarsely decomposed as follows:

$$
\begin{aligned}
& P(A^1, B^1, A^2, B^2, A^3, B^3) \qquad (8)\\
&= P(A^3|A^2)P(B^3|A^3)P(A^2|A^1)P(B^2|A^2)P(A^1)P(B^1|A^1)
\end{aligned}
$$

Based on the component subgraphs, we use the Theorem 2 and Lemma 4 in [4] to further decompose each term into the product of potential functions or conditional probabilities. The conditional probability $P(B^l|A^l)$ can be similarly factorized as the single-scale hybrid model in Eq.(5). The conditional probability $P(A^{l+1}|A^l)$ is actually the transition function between the adjacent layers. It has also been used in the multiscale MRF models (eg. [3, 27]). Our definition of $P(A^{l+1}|A^l)$ is as follows:

$$
\begin{aligned}
& P(A^{l+1}|A^l) \qquad (9)\\
&= P(\mathbf{y}^{(l+1)}, \mathbf{x}^{(l+1)}|\mathbf{y}^{(l)}, \mathbf{x}^{(l)})\\
&= P(\mathbf{y}^{(l+1)}, \mathbf{x}^{(l+1)}|\mathbf{y}^{(l)})\\
&\propto \prod_{y_j\in L_{l+1}}\phi(y_j, x_j^{(l+1)})\prod_{<j,k>\in L_{l+1}}\exp(y_jy_k\lambda^T g_{jk}(\mathbf{x}^{(l+1)}))\cdot\\
&\quad \prod_{y_s\in L_{l+1}, pa(y_s)\in L_l}\psi(y_s, pa(y_s))
\end{aligned}
$$

where $pa(y_s)$ denotes the parents of y_s at the next finer layer. The second equation is due to the Global Markov Property because $\{\mathbf{y}^{(l+1)}, \mathbf{x}^{(l+1)}\}$ is conditionally independent of $\mathbf{x}^{(l)}$ given $\mathbf{y}^{(l)}$. The term $\phi(y_j, x_j^{(l+1)})$ is the unary potential. The term $\exp(y_jy_k\lambda^T g_{jk}(\mathbf{x}^{(l+1)}))$ is the pairwise potential. The term $\psi(y_s, pa(y_s))$ is defined according to the causal inter-layer relationships between region nodes. For example, if 60% pixels of the region y_s are classified as the label $+1$ at the next finer layer given the configuration of $pa(y_s)$, then $\psi(y_s = 1, pa(y_s))$ is set as 60%. In other configurations, $\psi(y_s, pa(y_s))$ are similarly defined.

For a K-layer multiscale hybrid model, the JPD is factorized as follows:

$$
\begin{aligned}
& P(\{A^l, B^l\}_{l=1}^K) \qquad (10)\\
&= P(A^1)P(B^1|A^1)\prod_{l=1}^{K-1}P(A^{l+1}|A^l)P(B^{l+1}|A^{l+1})\\
&= \frac{1}{\mathbb{Z}}\prod_{y_j\in L_1}\phi(y_j, x_j^{(1)})\prod_{\substack{<j,k>\in L_1\\ k\in\mathcal{N}_j}}\exp(y_jy_k\lambda^T g_{jk}(\mathbf{x}^{(1)}))\cdot P(B^1|A^1)\cdot\\
&\quad \prod_{l=1}^{K-1}[\prod_{y_j\in L_{l+1}}\phi(y_j, x_j^{(l+1)})\cdot\prod_{\substack{<j,k>\in L_{l+1}\\ k\in\mathcal{N}_j}}\exp(y_jy_k\lambda^T g_{jk}(\mathbf{x}^{(l+1)}))\cdot\\
&\quad \prod_{y_s\in L_{l+1}, pa(y_s)\in L_l}\psi(y_s, pa(y_s))\cdot P(B^{l+1}|A^{l+1})\,]
\end{aligned}
$$

where $\mathbb{Z}$ is the normalization constant. $\mathcal{N}_j$ is the neighborhood of the jth region at the $(l+1)$th layer. The terms $P(B^1|A^1)$ and $P(B^{l+1}|A^{l+1})$ can be factorized according to Eq.(5).

Substituting Eq.(5) into Eq.(10), the JPD in the multiscale hybrid model is finally factorized as follows:

$$
\begin{aligned}
& P(\{A^l, B^l\}_{l=1}^K) \qquad (11)\\
&= P(A^1)P(B^1|A^1)\prod_{l=1}^{K-1}P(A^{l+1}|A^l)P(B^{l+1}|A^{l+1})\\
&= \frac{1}{\mathbb{Z}}\prod_{y_j\in L_1}\phi(y_j, x_j^{(1)})\prod_{<j,k>\in L_1, k\in\mathcal{N}_j}\exp(y_jy_k\lambda^T g_{jk}(\mathbf{x}^{(1)}))\ \bullet\\
&\quad \prod_{l=1}^{K-1}[\prod_{y_j\in L_{l+1}}\phi(y_j, x_j^{(l+1)})\prod_{\substack{<j,k>\in L_{l+1}\\ k\in\mathcal{N}_j}}\exp(y_jy_k\lambda^T g_{jk}(\mathbf{x}^{(l+1)}))\\
&\quad \prod_{y_s\in L_{l+1}, pa(y_s)\in L_l}\psi(y_s, pa(y_s))]\ \bullet\\
&\quad \prod_{l=1}^{K}[\prod_{e_j\in C^l}P(M_{e_j}|e_j)\prod_{v_t\in D^l}P(v_t|pa(v_t))P(M_{v_t}|v_t)\\
&\quad \prod_{e_j\in C^l}P(e_j|pa(e_j))\prod_{<j,k>,(e_j,e_k)\in C^l}h(e_j,e_k)]
\end{aligned}
$$

4.3. Parameter Learning

The parameter setting in the multiscale hybrid model is not a trivial task. However, we can decompose the parameter learning due to the conditional independence encoded in the graphical structure. For the child nodes in the directed part of the hybrid model, they follow the same conditional

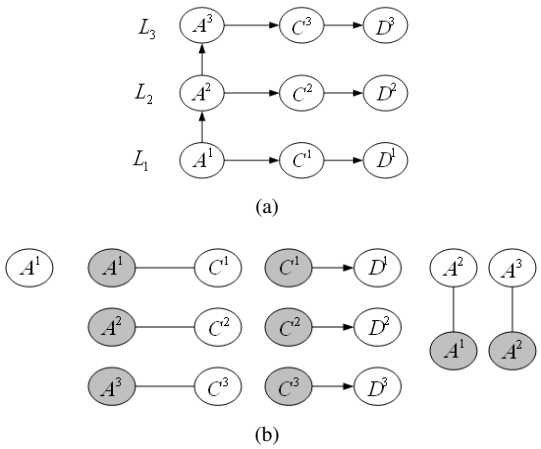


Figure 6. (a) the master graph of the multiscale hybrid model. (b) the component subgraphs of the multiscale hybrid model.

independence rules analogous to the d-separation rules [23] in a standard Bayesian Network. This property can simplify the parameter learning of the conditional probability distributions (CPDs). Given the complete training data (i.e. all labels are given), the CPDs of each child node can be independently learned with the labels of the child node and of its parents. For example, we can separately learn the likelihood model of edge measurements by Mixture of Gaussian analysis. The learning of CPDs for discrete nodes is simplified as counting the frequency of certain joint parent-child configurations in the training data.

The parameter learning for the undirected links is more difficult. We first learn the unary potential ϕ by training the MLP classifier. We then learn the weight λ of the pairwise potential by Maximum Likelihood Estimation (MLE). We can prove that the partial derivative of the log-likelihood w.r.t the weight λ is proportional to the difference between the empirical distribution and the expected distribution of the pairwise region labels. This result is due to the use of a log-linear model to define the pairwise potentials. It is similar as the standard MLE parameter learning in a CRF model. The main difference is that the expected distribution should be estimated directly from the hybrid model, which is more complex than the inference in a CRF model. Due to the space limitation, we omit the full derivations here.

5. Factor Graph Inference

The hybrid graphical model consists of both the directed links and the undirected links. To perform a consistent inference, it is necessary to convert the hybrid model into a Factor Graph (FG) representation [14, 8] since it is difficult to directly perform inference in such a hybrid model. A factor graph is a bipartite graph that expresses the structure of the factorization of a global function over a set of variables. The FG consists of two types of nodes: the variable nodes and the factor nodes. The variable node corresponds to a random variable, while the factor node represents the factorized local function. There is an edge connecting a variable node to a factor node if and only if the variable is an argument of the factorized function.

Since we already factorize the JPD of the proposed hybrid model (Eq.(11)), we can easily convert it into a factor graph representation, following the rules in [8]. Given the factor graph, there are different ways to perform probabilistic inference. Besides the sum-product and max-product algorithm, there are other algorithms that also can solve the inference problem. The stochastic local search (SLS) [22] is one of such algorithms. In [13], Hutter *et al*. improve SLS to achieve a more efficient algorithm for solving Most Probable Explanation (MPE) inference problem. Given the FG model, we use the inference package provided by Hutter *et al*. to perform MPE inference in the factor graph, i.e.

$$\{y^*, e^*, v^*\}_{l=1}^K = \arg \max_{\{y,e,v\}_{l=1}^K} P(\{y^l, x^l, e^l, v^l, M_e^l, M_v^l\}_{l=1}^K) \quad (12)$$

where the JPD is calculated by Eq.(11). In the MPE solution, the region nodes with the foreground labels at the finest layer yield the final segmentation.

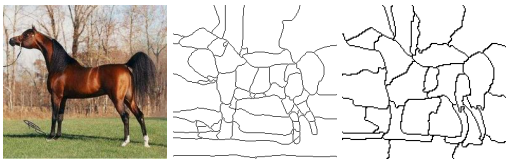
6. Experiments

We have tested the proposed multiscale hybrid graphical model on the Weizmann horse dataset [1] and the VOC2006 cow images [6]. The Weizmann horse dataset includes many horses that have different appearances, poses, as well as complex backgrounds. Several related works [5] [19] [28] also did experiments on this dataset. We can therefore compare our results with these state-of-the-art works.

We use 60 horse images as the training data to learn the model parameters. The testing images include 120 images from the Weizmann horse dataset. Compared to the training images, the foreground horses and the background scenes in the test images are more complex. The appearances of the horses have a much larger range of variations. The background includes more different kinds of scenes, some of which have never been seen in the training set.

We use the average CIELAB color features and their standard deviations as the local features x_i for each region. In this case, the three-layer perceptron has a structure with 6 nodes in the input layer, 35 nodes in the hidden layer and 1 node in the output layer.

For simplicity, we use a two-layer multiscale hybrid model in our experiments. All the training images and the test images are first oversegmented. Our framework does not require a specific algorithm for the oversegmentation. To demonstrate this, we use the Edgeflow-based Anisotropic diffusion method [26] to oversegment the image at the fine scale. The original image is then downsampled to 60% to generate the coarse level image. We use a



(a) (b) (c)

Figure 7. An example image and its oversegmentation. a)the color image; b)oversegmentation by Anisotropic diffusion at the fine scale; c)oversegmentation by Normalized Cut at the coarse scale.

Figure 8. Examples of the color image segmentation results produced by the proposed method, arranged in 2 groups of 2 rows. In each group, the first row includes the color images. The second row includes the segmentation masks produced by the multiscale hybrid graphical model.

standard Normalized Cut [25] to segment the coarse image. Figure 7 shows the oversegmentation at the fine scale and at the coarse scale by different methods. Given the training images and their ground truth labeling, we automatically train the hybrid graphical model.

After training the model, we perform image segmentation on the test images using the inference process described in section 5. Figure 8 shows examples of the color horse images and their segmentation masks. We achieved encouraging results on these images. Most small errors happen on the horse feet where the appearances of these parts are different from the horse body.

In order to quantitatively evaluate our segmentation results and compare with those aforementioned approaches, we calculate the average percentage of correctly labeled pixels (i.e. segmentation accuracy) in all test images. The quantitative results of our experiments are listed in Table 1. From the quantitative results in Table 1, we conclude that our results are better than the results produced by other state-of-the-art approaches. Note that we have not performed the feature selection like the work [19] has done. Besides, we have not utilized the additional object shape information as some works have exploited [5]. In Table 1, we also list the performance using a simple CRF model (Eq.(6)) as a baseline model for comparison. Apparently, its performance is inferior to the multiscale hybrid graphical model because the latter exploits more useful contextual relationships and information.

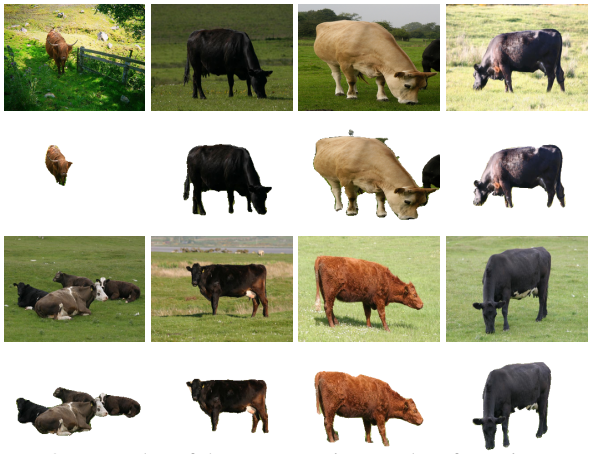


Figure 9. Examples of the segmentation results of cow images arranged in 2 groups of 2 rows. In each group, the first row includes the color images. The second row includes the segmentation masks produced by the multiscale hybrid graphical model.

Table 1. The quantitative comparison of our approach with several related works for segmenting the Weizmann horse images. The average percentage of correctly labeled pixels (i.e. segmentation accuracy) is used as the quantitative measurement.

method	image type	segmentation accuracy
Cour *et al*. [5]	color	94.2%
Levin *et al*. [19]	color	95%
Winn *et al*. [28]	color	93.1%
our multiscale hybrid model	color	96%
our CRF model alone	color	92.5%

We have also performed the experiments on a set of cow images from the VOC2006 database [6]. This database is primarily used for object categorization. In this work, we use it to test our image segmentation framework. Since there are no original ground-truth segmentations, we manually segment a set of cow images from this database. We use about a half set of the images (57 images) for training our model and use the rest half set of images (51 images) for testing. Figure 9 shows some examples of the image segmentation results. We have achieved good segmentation results on them. Although those cows have different appearances and sizes and there might be multiple cows in the image, our approach successfully segment them out. Besides the qualitative inspection of these results, we use the manual segmentation as the ground truth to calculate the segmentation accuracy. We achieve a good segmentation accuracy of 96.7% on these cow images.

We implement the whole model using Matlab software. The average size of the test horse images is 255×207 pixels. The average size of the cow images is 256×192 pixels. The segmentation speed mainly depends on the complexity of the constructed graphical model. It typically takes about 10 to 30 seconds to segment one image using the efficient

Factor Graph inference [13] in a Pentium 1.7GHz laptop.

7. Conclusions

In this paper, we present a multiscale hybrid model that can systematically model heterogeneous contextual relationships between random variables. We use this framework to capture the natural causalities between multiple image entities, the spatial relationships between region nodes, the inter-layer consistency between region nodes at different scales, and the smoothness constraint between edges, etc.

Based on the advanced graphical model theory, we factorize the joint probability distribution of the hybrid model into the products of conditional probabilities and potential functions. With this factorization, we convert the hybrid model into a factor graph representation to perform joint probabilistic inference in a principled way. Our experiments on the figure/ground image segmentation problems demonstrate the usefulness of the proposed framework for effective and robust image segmentation. Moreover, the application of the proposed framework is not limited to image segmentation. In fact, it can be applied to many computer vision problems where heterogeneous contextual information is important, including object tracking, object recognition, activity recognition, etc. We will explore the applications of this model to other computer vision problems in future.

References

[1] E. Borenstein, E. Sharon, and S. Ullman. Combining top-down and bottom-up segmentation. In *CVPR Workshop on Perceptual Organization in Computer Vision*, pages 46– 46, 2004.

[2] C. Bouman and B. Liu. Multiple resolution segmentation of textured images. *PAMI*, 13(2):99–113, 1991.

[3] C. Bouman and M. Shapiro. A multiscale random field model for bayesian image segmentation. *IEEE Trans. on Image Processing*, 3(2):162–177, March 1994.

[4] W. L. Buntine. Chain graphs for learning. In *Conference on Uncertainty in Artificial Intelligence*, pages 46–54, 1995.

[5] T. Cour and J. Shi. Recognizing objects by piecing together the segmentation puzzle. In *CVPR*, pages 1–8, 2007.

[6] M. Everingham, A. Zisserman, C. Williams, and L. V. Gool. The pascal visual object classes challenge 2006 (voc 2006) results. Technical report, Oxford, UK., 2006.

[7] X. Feng, C. Williams, and S. Felderhof. Combining belief networks and neural networks for scene segmentation. *PAMI*, 24(4):467–483, 2002.

[8] B. J. Frey. Extending factor graphs so as to unify directed and undirected graphical models. In *Conference on Uncertainty in Artificial Intelligence 19*, pages 257–26, 2003.

[9] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *PAMI*, 6(6):721–741, 1984.

[10] C. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey Vision Conference*, pages 147 – 152, 1988.

[11] X. He, R. Zemel, and M. Carreira Perpinan. Multiscale conditional random fields for image labeling. In *CVPR*, volume 2, pages 695–702, 2004.

[12] X. He, R. S. Zemel, and D. Ray. Learning and incorporating top-down cues in image segmentation. In *ECCV*, pages 338–351, 2006.

[13] F. Hutter, H. H. Hoos, and T. Stutzle. Efficient stochastic local search for mpe solving. In *IJCAI*, pages 169–174, 2005.

[14] F. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.

[15] S. Kumar and M. Hebert. Discriminative random fields. *IJCV*, 68(2):179–201, 2006.

[16] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, pages 282C–289, 2001.

[17] S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

[18] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *CVPR*, pages 316–323, 2000.

[19] A. Levin and Y. Weiss. Learning to combine bottom-up and top-down. segmentation. In *ECCV*, pages 581–594, 2006.

[20] D. Melas and S. Wilson. Double markov random fields and bayesian image segmentation. *IEEE Trans. on Signal Processing*, 50(2):357–365, 2002.

[21] M.Kass, A.Witkin, and D.Terzopoulos. Snakes: Active contour models. *IJCV*, 1:321–331, 1988.

[22] J. Park. Using weighted max-sat engines to solve mpe. In *Proceedings of the 18th National Conference on Artificial Intelligence (AAAI)*, pages 682–687, 2002.

[23] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan-Kaufmann Publishers Inc., 1988.

[24] S. Sarkar and K. L. Boyer. Integration, inference, and management of spatial information using bayesian networks: Perceptual organization. *PAMI*, 15(3):256–274, 1993.

[25] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 22(8):888–905, 2000.

[26] B. Sumengen and B. S. Manjunath. Edgeflow-driven variational image segmentation: Theory and performance evaluation. Technical report, University of California, Santa Barbara, 2005. http://barissumengen.com/seg/.

[27] R. Wilson and C.-T. Li. A class of discrete multiresolution random fields and its application to image segmentation. *PAMI*, 25(1):42– 56, 2002.

[28] J. Winn and N. Jojic. Locus: Learning object classes with unsupervised segmentation. In *ICCV*, pages 756–763, 2005.

[29] S. C. Zhu and A. Yuille. Region competition: Unifying snake/balloon, region growing and bayes/mdl/energy for multi-band image segmentation. *PAMI*, 18(9):884–900, 1996.