Image Segmentation with a Unified Graphical Model

Lei Zhang, *Member*, *IEEE*, and Qiang Ji, *Senior Member*, *IEEE*

Abstract—We propose a unified graphical model that can represent both the causal and noncausal relationships among random variables and apply it to the image segmentation problem. Specifically, we first propose to employ Conditional Random Field (CRF) to model the spatial relationships among image superpixel regions and their measurements. We then introduce a multilayer Bayesian Network (BN) to model the causal dependencies that naturally exist among different image entities, including image regions, edges, and vertices. The CRF model and the BN model are then systematically and seamlessly combined through the theories of Factor Graph to form a unified probabilistic graphical model that captures the complex relationships among different image entities. Using the unified graphical model, image segmentation can be performed through a principled probabilistic inference. Experimental results on the Weizmann horse data set, on the VOC2006 cow data set, and on the MSRC2 multiclass data set demonstrate that our approach achieves favorable results compared to state-of-the-art approaches as well as those that use either the BN model or CRF model alone.

Index Terms—Image segmentation, probabilistic graphical model, Conditional Random Field, Bayesian Network, factor graph.

1 Introduction

Image segmentation has been an active area of research in computer vision for more than 30 years. Many approaches have been proposed to solve this problem. They can be roughly divided into two groups: the deterministic approach and the probabilistic approach. The former formulates the segmentation problem as a deterministic optimization problem. This approach includes the clustering method [1], "snakes" or active contours [2], the graph partitioning method [3], the level set-based method [4], etc. The probabilistic approach, on the other hand, formulates the segmentation problem as a stochastic optimization problem. It can be further divided into two groups. One group uses various graphical models (such as Markov Random Fields and Bayesian Network) to model the joint probability distribution of the related image entities [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. The other group directly models the probability distribution of the image entities either parametrically or nonparametrically, without using graphical models. It includes the discriminative approach [16], [17], [18], [19], the generative approach [20], [21], [22], [23], and the hybrid approach, combining the discriminative model and the generative model [24], [25]. Our work belongs to the category of using graphical models for image segmentation. Specifically, we develop a unified graphical model that can incorporate various types of probabilistic relationships and apply it to the image segmentation problem. However, this framework is general enough to be applied to other computer vision problems.

Much progress has been made in the image segmentation field so far. As a result of the progress, computer vision is able to segment increasingly more complex images. Image segmentation, however, is still far from being resolved. One factor that prevents this from happening is the lack of ability by the existing methods to incorporate information/knowledge other than image data itself. Many existing image segmentation methods are data driven. These methods tend to fail when image contrast is low or in the presence of occlusion, the clutter of other objects. The fact of the matter is that the image itself may not contain enough information for an effective segmentation, no matter what algorithms we use and how sophisticated the algorithms are.

If we study human segmentation, we will quickly realize that humans tend to exploit additional knowledge besides the image data to perform this task. Humans segment an image not only based on the image itself, but also based on their plentiful knowledge, such as the contour smoothness, connectivity, local homogeneity, the object shape, the contextual information, etc. A human's capability to combine image data with additional knowledge plays an important role for effective and robust image segmentation. Many researchers have realized this aspect and have proposed different model-based approaches for image segmentation. The model is used to capture certain prior knowledge and to guide the image segmentation.

Despite these efforts, what is lacking is an image segmentation model that can systematically integrate different types of prior knowledge and the image data. Many existing approaches can only exploit very limited information, such as the image data and the local homogeneity of image labels. One of the reasons is due to the lack of a systematic way to incorporate various types of knowledge into a single framework.

A desirable image segmentation framework may be the one that is able to flexibly incorporate various types of information and constraints, and solve image segmentation in a probabilistic way. We notice that Probabilistic Graphical Models (PGMs) [26], [27], [28], [29], [30] in the Machine

• *The authors are with Rensselaer Polytechnic Institute, Troy, NY 12180. E-mail: leizhang2009@gmail.com, qji@ecse.rpi.edu.*

Manuscript received 2 Dec. 2008; revised 5 Mar. 2009; accepted 18 May 2009; published online 9 July 2009.
Recommended for acceptance by F. Dellaert.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2008-12-0828.
Digital Object Identifier no. 10.1109/TPAMI.2009.145.

0162-8828/10/$26.00 © 2010 IEEE Published by the IEEE Computer Society

Learning community are very powerful statistical models that are potentially able to satisfy all of these requirements. They provide an effective way to model various types of image entities, their uncertainties, and the related prior knowledge.

There are two basic types of graphical models: the undirected graphical model and the directed acyclic graphical model. The undirected graphical model can represent noncausal relationships among the random variables. The Markov Random Field (MRF) [5], [26] is a type of well-studied undirected graphical model. MRF models have been widely used for image segmentation. They incorporate the spatial relationships among neighboring labels as a Markovian prior. This prior can encourage (or discourage) the adjacent pixels to be classified into the same group. As an extension to MRFs, the Conditional Random Field (CRF) [27] is another type of undirected graphical model that has become increasingly popular. The differences between MRF models and CRF models will be elaborated in Section 2.

While both MRF and CRF models can effectively capture noncausal relationships among the random variables (i.e., the nodes in a graphical model), such as the spatial homogeneity, they cannot model some directed relationships (e.g., the causalities) that extensively exist and are also important [31]. Fortunately, this problem can be complementarily solved by another type of graphical model, i.e., the directed acyclic graphical model such as Bayesian Network (BN) [29] [30]. BN can conveniently model the causal relationships between random variables using directed links and conditional probabilities. It has been successfully applied to medical diagnosis systems, expert systems, decision-making systems, etc. For image segmentation, there are some relationships that can be naturally modeled as causal relationships. For example, two adjacent regions with significantly different characteristics can lead to a high-contrast edge between them. In another example, the mutual exclusion, co-occurrence, or intercompetition relationships among the intersecting edges can also be modeled as causal relationships. These relationships are useful when one wants to impose some constraints on the edges. For example, two adjacent edges initially might have no relationships with each other when deciding which edge is part of the object boundary. However, knowing one edge is part of the object boundary and that the object boundary should be smooth, the probability of the other edge on the object boundary will reduce if the angle between the two edges is small. In this case, two edges become dependent on each other now. Such a relationship can be modeled by BN as the "explaining-away" relationship, but is hard to model by the undirected graphical models.

The existing graphical models for image segmentation tend to be either directed or undirected models alone. While they can effectively capture one type of image relationship, they often fail to capture the complex image relationships of different types. To overcome this limitation, we propose a probabilistic framework that unifies an undirected graphical model (i.e., the CRF model) with a directed graphical model (i.e., the BN model). It can flexibly incorporate image measurements, both noncausal relationships and causal relationships, and various types (both quantitative and qualitative) of human prior knowledge. In addition, the unified model systematically integrates the region-based image segmentation with the edge-based image segmentation. With this framework, image segmentation is performed through a principled probabilistic inference. The proposed framework is powerful, flexible, and also extendable to other computer vision problems.

Our main contributions lie in the introduction of a unified probabilistic framework for effective and robust image segmentation by incorporating various types of contextual/prior knowledge and image measurements under uncertainties. Our model captures the natural causal relationships among three entities in image segmentation: the regions, edges, and vertices (i.e., the junctions) as well as the noncausal spatial relationships among image labels and their measurements. Besides, various constraints are also modeled as either directed relationships or undirected relationships in the model.

The remainder of this paper is organized as follows: In Section 2, we review the related works that use graphical models for image segmentation. In Section 3, we give an overview of the proposed unified graphical model. In Section 4, we describe the region-based CRF image segmentation model. In Section 5, we describe the edge-based BN image segmentation model. In Section 6, we explain how we combine the CRF model with the BN model into a unified graphical model. In Section 7, we introduce the experiments on the Weizmann horse data set [32], on the Microsoft multiclass segmentation data set (MSRC2) [33], and the comparisons with the state-of-the-art techniques. In addition, we also introduce the experiments on the cow images from the VOC2006 database [34]. This paper concludes in Section 8.

2 Related Work

Various graphical models have been used for image segmentation. In particular, the Markov Random Field has been used for a long time [5], [6], [7], [8], [9]. In the simplest case, the MRF model formulates the joint probability distribution of the image observation and the label random variables on the 2D regular lattice. It is therefore a generative model. According to the Bayes' rule, the joint probability can be decomposed into the product of the likelihood of the image observation and the prior distribution of the label random variables. An apriori Markovian field is normally assumed as the prior distribution, which normally encourages the adjacent labels to be the same (i.e., locally homogeneous). In order to reduce the computational complexity, the MRF model often assumes the observations to be conditionally independent given the label of each site.

The simple MRF model has been extended to more complex structures. In [6], the authors propose a multiscale hierarchical model. The label random variables at two adjacent layers form a Markov chain. The links between them enforce the consistency of the labels at adjacent layers. This model can partially model the long-range relationships between spatially faraway nodes. One problem of this model is due to its model structure. The commonly used quadtree model leads to the problem that spatially adjacent nodes may be far from each other in the quadtree structure. Another problem is that there are no direct interactions

among adjacent labels at the same layer. These issues can lead to imprecise segmentation at the boundary [35].

Several works extend the idea of multiscale random fields to make it more powerful. Cheng and Bouman [36] extend it by introducing a more complex transition conditional probability using a class probability tree. The advantage of using such an approach is that it can use a relatively larger size (e.g., 5×5) of the neighborhood at the next coarser layer. By considering more parents, complex context information can be taken into account in the multiscale random fields. Wilson and Li [37] also try to improve the interactions of the label fields at two adjacent layers. The neighborhood of a site is extended to include two types of sites: the parent sites at the next coarser layer and the adjacent sites at the same layer. In [38], Irving et al. alleviate the drawback of the quadtree model due to its fixed structure. Instead of assuming the nodes at one layer to be nonoverlapped, they propose an overlapping tree model where the sites at each layer correspond to overlapping parts in the image. This approach is also demonstrated to reduce the problem of imprecise segmentation.

The conditional independence assumption of the image observations in a MRF model is normally invalid for texture image segmentation. A double Markov Random Fields (DMRF) model is exploited [7] in order to overcome this limitation. In this model, one MRF is used to represent the labeling process and another MRF is used to model the textured regions. With the DMRF model, the textured image segmentation can be performed by simultaneously estimating the MRF texture models and the MRF label fields. In [8], the authors propose a pairwise Markov Random Fields (PMF) to overcome the strong conditional independence assumption. They relax this assumption by directly assuming the joint random fields of the labels and the observations follow the Markovian property. Although the PMF model has weaker assumptions than the traditional MRF model, the authors shift the problem to directly model the joint probability, which is generally a difficult problem.

Differently from the MRF model, the CRF [27] directly models the posteriori probability distribution of the label random variables, given the image observation. This posteriori probability is assumed to follow the Markovian property. The CRF is therefore a discriminative model that focuses on discriminating image observations at different sites. Since the CRF model does not try to describe the probability distribution of the observation, it may require fewer resources for training, as pointed out by [27], [39]. Compared to the traditional MRF model, the CRF model relaxes the conditional independence assumption of the observations. The CRF model allows arbitrary relationships among the observations, which is obviously more natural in reality. The CRF model also makes the Markovian assumption of the labels conditioned on the observation. As a result, the CRF model can relax the apriori homogeneity constraint based on the observation. For example, this constraint may not be enforced where there is a strong edge. This characteristic makes the CRF model able to handle the discontinuity of the image data and labels in a natural way.

Several previous works have demonstrated the success of CRF models in image segmentation. He et al. [10] have used CRF for segmenting static images. By introducing the hidden random variables, they can incorporate additional contextual knowledge (i.e., the scene context) to facilitate image segmentation. In [11], Ren et al. have used CRF for figure/ground labeling. They first oversegment the image into a set of triangles using the constrained Delaunay triangulation (CDT). A CRF model is then constructed based on these triangles. Their CRF model integrates various cues (e.g., similarity, continuity, and familiarity) for image segmentation by adding additional energy functions. This model, however, treats all hidden random variables as in the same layer and ignores the fact that they may come from different levels of abstraction. Our model is different from Ren et al.'s model because we model different image entities at the hierarchical layers and capture their natural causal relationships within this hierarchical framework.

Besides the simple CRF model, more complex CRF models have also been proposed. A hierarchical CRF model was proposed by Kumar and Hebert to exploit different levels of contextual information for object detection [40]. This model can capture both pixelwise spatial interactions and relative configurations between objects. Another hierarchical tree-structured CRF model was developed in [12]. These models are in spirit similar to the Multiscale Random Field model in [35]. The main difference is that all links are now represented by the undirected links and the pairwise relationships are modeled by the potential functions conditioned on the image observation. In [41], the authors developed a complex CRF model for object detection. They introduced additional hidden layers to represent the locations of the detected object parts. The object parts are constrained by their relative locations with respect to the object center. Based on a model similar to [41], Winn and Shotton further introduced the layout consistency relationships among parts for recognizing and segmenting partially occluded objects in both 2D [42] and 3D [43] cases.

Although not as popular as those undirected graphical models (MRF or CRF), directed graphical models such as the Bayesian Network (BN) have also been exploited in solving computer vision problems [44], [45], [13], [14], [15]. BN provides a systematic way to model the causal relationships among the random variables. It simplifies the modeling of a possibly complex joint probability distribution by explicitly exploiting the conditional independence relationships (known as prior knowledge) encoded in the BN structure. Based on the BN structure, the joint probability is decomposed into the product of a set of local conditional probabilities, which is much easier to specify because of their semantic meanings.

For image segmentation, the BN model can be used to represent the prior knowledge of the statistical relationships among different entities, such as the regions, edges, and their measurements. Several previous works have exploited BN for image segmentation. Feng et al. [13] combine BN with Neural Networks (NN) for scene segmentation. They train neural networks as the classifiers that can produce the "scaled-likelihood" of the data. The BN models the prior distribution of the label fields. The local predictions of pixelwise labels generated from NN are fused with the prior to form a hybrid segmentation approach. Since Feng's BN model has the quadtree structure, it inherits the drawback of

the quadtree model due to its fixed structure. In order to overcome such a problem, Todorovic and Nechyba [46] develop a dynamic multiscale tree model that simultaneously infers the optimal structure as well as the random variable states. Although they show some successful experiments, their model is very complex and there are many random variables to be inferred. Good initialization is required for their variational inference approach. In [14], Mortensen and Jia proposed a semiautomatic segmentation technique based on a two-layer BN. Given a user-input seed path, they use the minimum-path spanning tree graph search to find the most likely object boundaries. In [15], Alvarado et al. use a BN model to capture all available knowledge about the real composition of a scene for segmenting a handheld object. Their BN model combines high-level cues such as the possible locations of the hand in the image to infer the probability of a region belonging to the object.

Besides those segmentation approaches based on either an undirected graphical model or a directed graphical model, several hybrid approaches have been previously proposed to combine different types of segmentation techniques. Huang et al. [47] couple an MRF with a deformable model for image segmentation. To make the inference tractable, they require decoupling the model into two separate parts and using different techniques to perform inference in each part. The inference in the MRF part is performed using belief propagation, while the estimation of the deformable contour is based on the variational approaches. In contrast, our model unifies two types of graphical models (i.e., CRF and BN) and belief propagation can be applied to both models. Therefore, we can use a consistent inference approach in our model. Lee et al. [48] combine Discriminative Random Field (DRF) [39] with Support Vector Machines (SVMs) to segment brain tumors. They use the SVM to train a classifier that predicts a label based on the local feature. The local prediction is then combined with a prior distribution modeled by a DRF model for image labeling. In this approach, the two model parts (i.e., SVM and DRF) are not functioning at the same level. The DRF serves as the backbone of the whole model, while the SVM serves as the local classifier to collect image features for labeling. Moreover, the two parts are operated under different principles and cannot be unified in a consistent way.

As discussed before, the undirected graphical model (e.g., MRF and CRF) and the directed graphical model (e.g., BN) are suitable for representing different types of statistical relationships among the random variables. Their combination can create a more powerful and flexible probabilistic graphical model that can easily model various apriori knowledge to facilitate image segmentation. This is the basic motivation of our work in this paper.

Little previous work focuses on modeling image segmentation by unifying different types of graphical models. Liu et al. [49] combine a BN with a MRF for image segmentation. A naive BN is used to transform the image features into a probability map in the image domain. The MRF enforces the spatial relationships of the labels. The use of a naive BN greatly limits the capability of this method because it is hard to model the complex relationships between the label random variables and the image measurements using a naive BN. In contrast, we use a hierarchical BN to capture the complex causalities among multiple image entities and their measurements, as well as the local constraints. Murino et al. [50] formulate a Bayesian Network of Markov Random Field model for image processing. A BN is used to represent the apriori constraints between different abstraction levels. A coupled MRF is used to solve the coupled restoration and segmentation problem at each level. This approach only uses the BN to model the set of apriori constraints between the same entities at different levels. The image information is only exploited in the MRF model for inferring the hidden random variables, while, in our unified model, we exploit image measurements both in the CRF part and in the BN part to perform image segmentation based on two complementary principles: the region-based segmentation and the edge-based segmentation. In [51], Kumar et al. combined an MRF with a layered pictorial structures (LPS) model for object detection and segmentation. The LPS model represents the global shape of the object and restrains the relative location of different parts of the object. They formulate the LPS model using a fully connected MRF. The whole model is therefore an extended MRF model, which is different from our unified model which combines different types of graphical models. Hinton et al. [52] studied the learning issue for a hybrid model. Their hybrid model differs from ours in several aspects. First, Hinton et al.'s model is constructed by connecting several MRFs at different layers using directed links. The configuration of a top-level MRF provides the biases that influence the configuration of the next level MRF through the directed links, while, in our model, the directed links capture the causalities among multiple image entities and the undirected links capture the spatial correlation conditioned on the observation. Second and most important, Hinton et al. exploit an approximation of the true posterior probability distribution of the hidden nodes by implicitly assuming the posterior of each hidden node is independent of each other. In contrast, we derive the factored joint probability distribution using the global Markov property based on the graphical model structure, and therefore, do not have such an assumption as Hinton et al.'s. Third, based on their approximation, Hinton et al. apply variational approach to approximately perform inference and parameter learning. In contrast, we convert our model into a factor graph to perform inference as well as learning through principled factor graph inference.

Compared to the aforementioned works, our unified model differs from them in several aspects. First, our model can capture both the causal and noncausal relationships that extensively and naturally exist among different image entities. Second, our unified model consists of two parts: a CRF part and a BN part. It can therefore systematically combine the region-based image segmentation (i.e., the CRF part) with the edge-based image segmentation (i.e., the BN part). Finally, via the factor graph, we can perform image segmentation under the unified framework through a consistent probabilistic inference.

3 Overview of the Approach

The structure of the proposed graphical model is illustrated in Fig. 1. It consists of two parts: the CRF part and the BN part. The CRF part in Fig. 1a performs region-based image segmentation, while the BN part performs edge-based

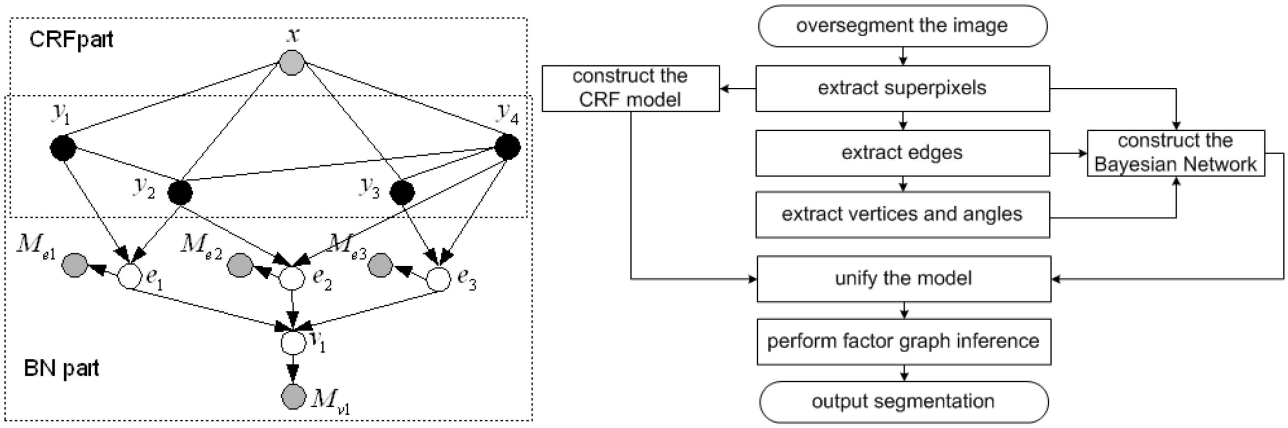

Fig. 1. (a) The graphical structure of the unified image segmentation model. It combines the Conditional Random Field with the Bayesian Network to unify the region-based image segmentation and the edge-based image segmentation. This example includes four superpixel region nodes $\{y_i\}_{i=1}^4$, three edge nodes $\{e_j\}_{j=1}^3$ and their measurement $\{M_{e_j}\}_{j=1}^3$, one vertex node v_1 and its measurement M_{v_1}. (b) The process to construct the CRF model and the BN model from the oversegmentation of the original image.

segmentation. The two parts are connected through the region nodes $\{y_i\}$.

Given an image, it is first oversegmented to produce an edge map. We can use any suitable segmenter to produce this oversegmentation. Specifically, we use the Edgeflow-based anisotropic diffusion method [53] for this purpose. The region node y_i correspond to each oversegmented region (referred to as superpixel thereafter) in the edge map. We assume that the superpixel region nodes $\mathbf{y} = \{y_i\}$ and the image observation $\mathbf{x}$ form a CRF. They are used to construct the CRF part of the model. Conditioned on the image observation $\mathbf{x}$, the region nodes $\mathbf{y}$ follow the Markovian property, i.e., $P(y_i|\{y_j\}_{j\neq i}, \mathbf{x}) = P(y_i|\mathcal{N}(y_i), \mathbf{x})$, where $\mathcal{N}(y_i)$ represents the spatial neighborhood of y_i.

The CRF image segmentation can be thought of as a labeling problem, i.e., to assign a label to the ith superpixel. In the figure/ground image segmentation problem, the node y_i has two possible labels, i.e., $+1$ (the foreground) and -1 (the background). Let $\{y_i\}_{i=1}^n$ be the label random variables corresponding to all superpixels, where n is the total number of superpixels in an image. $\{x_i\}_{i=1}^n$ are the corresponding image measurements. x_i is represented as a local feature vector extracted from the image. Different types of cues such as intensity, color, and textures can be included in this feature vector for image segmentation. Using the CRF model, we can infer the label for each superpixel from image measurements.

While the CRF model can effectively capture the spatial relationships among image entities, they cannot capture other causal relationships that naturally exist among different image entities. Due to this reason, a multilayer BN is constructed to capture the causalities among the regions, edges, vertices (or junctions), and their measurements. In Fig. 1a, the $\mathbf{e}$ nodes represent all the edge nodes in the constructed BN model. These edge nodes correspond to the edge segments in the edge map. The $\mathbf{v}$ nodes represent all the vertex nodes. They are automatically detected from the edge map (see more details in Section 5). The nodes $\mathbf{M_e}$ and $\mathbf{M_v}$ represent the measurements of edges and vertices, respectively. Given the BN model, the goal is to infer the edge labels from various image measurements.

Combining the CRF model with the BN model yields a unified probabilistic graphical model that captures both the causal and noncausal relationships, as shown in Fig. 1a. The unified graphical model is further converted into a Factor Graph representation for performing the joint inference of image labels. Based on the Factor Graph theory, principled algorithms such as the sum-product algorithm and the max-product algorithm can be used to perform consistent inference in the unified model. We therefore formulate a unified graphical model that can exploit both the region-based information and the edge-based information, and more importantly, the causal and noncausal relationships among random variables for image segmentation. Specifically, in Fig. 1a, the region nodes $\{y_i\}$ act as the parents of an edge node. The parents of the edge node correspond to the two regions that intersect to form this edge. The links between the parents and the child represent their causal relationships. If the parent region nodes have different labels, it is more likely that there is an object boundary (i.e., $e_j = 1$) between them. In this way, the proposed model systematically combines the CRF model and the BN model in a unified probabilistic framework.

4 Region-Based Image Segmentation Using Conditional Random Field

Our CRF model is a superpixel-based model. We choose the superpixel-based model because the relationship between two regions naturally provides the clue for inferring the state of the edge between them. In addition, the superpixel CRF model reduces the computational problem that is a common issue in the undirected graphical model. Using the CRF model, we want to decide if the superpixels representing the oversegmented regions should be merged.

As shown in Fig. 2, the image is first oversegmented into superpixel regions. Each superpixel region is a relatively homogenous region and it corresponds to one region node in the CRF model. Based on the oversegmentation, we automatically detect the topological relationships among these region nodes. The CRF model is then constructed based on these topological relationships.

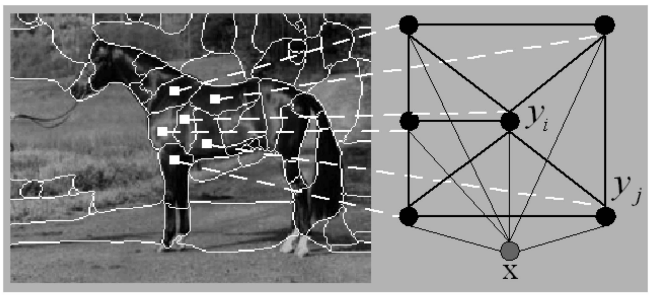

Fig. 2. Illustration of a part of the CRF model built on the superpixel regions. The oversegmentation is superimposed on the original image for an easy view. Each small region in the oversegmented image corresponds to a region node in the CRF model. The correspondence between the region nodes and the superpixel regions are indicated by the white dotted links. The black nodes are the region nodes. The gray node represents the whole image observation.

For the computational concern, we only consider the pairwise relationships among the region nodes. If two superpixel regions are adjacent to each other in the oversegmentation, an undirected link will be added between their corresponding nodes in the CRF model. This link means that there is an interaction between them, which is represented by the pairwise potential. From the example in Fig. 2, it is obvious that different region nodes may have a different number of neighbors, which means their interactions with other nodes can be stronger or weaker.

4.1 CRF Model

Our CRF model directly models the posteriori probability distribution of the region nodes $\mathbf{y}$ as

$$P(y|x) = \frac{1}{Z}\prod_{i\in V}\phi(y_i, x_i)\prod_{i\in V}\prod_{j\in\mathcal{N}_i} exp(y_i y_j \lambda^T g_{ij}(x)), \quad (1)$$

where V is the set of all superpixel region nodes and y is the joint labeling of all region nodes. $\mathcal{N}_i$ denotes the neighborhood of the ith region, which is automatically detected from the topological relationships. λ is the parameter vector. $g_{ij}(\cdot)$ represents the feature vector for a pair of nodes i and j. The symbol Z denotes the normalization term (i.e., the partition function).

There are two parts in (1). The first part, $\phi(y_i, x_i)$, is the unary potential, which tries to label the ith node according to its local features. It indicates how likely it is that the ith node will be assigned the label y_i given the local features x_i. For this purpose, we use a discriminative classifier based on a multilayer perceptron (MLP). We actually use a three-layer perceptron classifier in this work. Let $net(x_i)$ denotes the output of the perceptron when the feature vector x_i is the input. The output is further converted into a probabilistic interpretation using a logistic function,

$$\phi(y_i, x_i) = \frac{1}{1 + exp\big(-y_i \frac{net(x_i)}{\tau}\big)}, \quad (2)$$

where τ is a constant that can adjust the curve of the logistic function. Similar definitions of the unary potentials were first proposed in [54] and have been used in [10], [39]. The three-layer perceptron is automatically trained with a set of training data.

The second part $exp(y_i y_j \lambda^T g_{ij}(x))$ in (1) is the pairwise potential. It can be seen as a measure of how the adjacent region nodes y_i and y_j should interact with each other, given the measurements $\mathbf{x}$. We use a log-linear model to define this pairwise potential, which depends on the inner product of the weight vector λ and the pairwise feature vector $g_{ij}(x)$. The weight vector λ will be learned during a training process (see more details in Section 4.2). The pairwise feature vector $g_{ij}(x)$ can be defined based on the whole image measurements to consider the arbitrary relationships among the observations. For simplicity, it is currently defined based on the difference of the feature vectors x_i and x_j. An additional bias term (fixed as 1) is also added into the feature vector $g_{ij}(x)$. The feature vector $g_{ij}(x)$ is defined as $g_{ij}(x) = [1, |x_i - x_j|]^T$, where T is the transpose of a vector. The operator $|\cdot|$ represents the absolute value of each component in the difference between x_i and x_j. Similar pairwise potentials have been used in [39].

The pairwise potential $exp(y_i y_j \lambda^T g_{ij}(x))$ will have different values when the labels y_i and y_j are same or different. Moreover, when the measurements of the ith and jth nodes are same, the pairwise potential will only depend on the bias term. The bias term determines how the model prefers the neighboring nodes to have same or different labels. If the measurements at nodes i and j are significantly different, the pairwise potential will depend on the difference of measurements and the corresponding weights. This definition has incorporated the data adaptive capability, which is important for segmenting regions with discontinuity (e.g., the regions near strong edges).

The partition function Z in (1) can be calculated by summing out all the possible configurations of $\mathbf{y}$, i.e.,

$$Z = \sum_y exp\left\{\sum_{i\in V}\left[log\phi(y_i, x_i) + \sum_{j\in\mathcal{N}_i} y_i y_j \lambda^T g_{ij}(x)\right]\right\}. \quad (3)$$

Direct calculation of the partition function in (3) is computationally difficult. However, the Bethe free energy [55] can be used to approximate the logarithm of the partition function and to alleviate this computational problem.

4.2 Parameter Estimation

The three-layer perceptron classifier and the parameter vector λ of the pairwise potential are automatically learned from the training data. Assuming that we have $x^{(1)}, x^{(2)}, \ldots, x^{(K)}$ training images and their ground truth labeling $y^{(1)}, y^{(2)}, \ldots, y^{(K)}$, where K is the number of training images, the aim of parameter estimation is to automatically learn the three-layer preceptron classifier and the parameter vector λ from these data.

First, we train the three-layer perceptron classifier. The structure of our three-layer perceptron depends on the dimension of the input feature vector and the number of available training data. In the training stage, the target output of the three-layer perceptron is $+1$ (the foreground) or -1 (the background). Given the input and the desired output, the three-layer perceptron is trained using the standard BFGS quasi-Newton backpropagation method.

Next, we fix the three-layer perceptron classifier and use the conditional Maximum Likelihood Estimation (MLE) method to learn the parameter vector λ. Assuming all of the

training data are identically independently sampled, the log-likelihood of the parameter λ is calculated as

$$L(\lambda) = \sum_{k=1}^{K}\left\{\sum_{i\in V}\left[log\phi\left(y_i^{(k)}, x_i^{(k)}\right) + \sum_{j\in\mathcal{N}_i} y_i^{(k)} y_j^{(k)} \lambda^T g_{ij}\left(x^{(k)}\right)\right] - z^{(k)}\right\}, \tag{4}$$

where $z^{(k)}$ is the log-partition function, i.e., $z^{(k)} = logZ^{(k)}$. As mentioned before, this log-partition function can be approximated by the Bethe free energy.

The optimal parameters λ^* are estimated according to the MLE estimation, i.e.,

$$\lambda^* = \arg\max_{\lambda} L(\lambda) = \arg\min_{\lambda} -L(\lambda). \tag{5}$$

We use the stochastic gradient descent method [55] to find the optimal parameters λ^* in (5). The gradient of the log-likelihood $L(\lambda)$ is calculated as follows:

$$\frac{\partial L(\lambda)}{\partial \lambda} = \sum_{k=1}^{K}\left[\sum_i\sum_{j\in\mathcal{N}_i} y_i^{(k)} y_j^{(k)} g_{ij}\left(x^{(k)}\right) - E_{P(y|x^{(k)};\lambda)}\left(\sum_i\sum_{j\in\mathcal{N}_i} y_i y_j g_{ij}\left(x^{(k)}\right)\right)\right], \tag{6}$$

where $E_P[\cdot]$ denotes the expectation with respect to the distribution P. For example,

$$E_{P(y|x^{(k)};\lambda)}\left(\sum_i\sum_{j\in\mathcal{N}_i} y_i y_j g_{ij}\left(x^{(k)}\right)\right) = \sum_y P\left(y|x^{(k)};\lambda\right)\sum_i\sum_{j\in\mathcal{N}_i} y_i y_j g_{ij}\left(x^{(k)}\right). \tag{7}$$

The summation in (7) is performed over all possible configurations of the region nodes $\mathbf{y}$.

4.3 Labeling Inference

When all of the parameters are known, the CRF model can be used to infer the region labels corresponding to the superpixel regions. The optimal labeling can be found by the Maximum Posterior Marginal (MPM) criterion [56]. Each node i is assigned a label that maximizes its marginal posteriori probability, i.e.,

$$y_i^* = \arg\max_{y_i\in\{1,-1\}} P(y_i|x;\lambda). \tag{8}$$

The marginal probability $P(y_i|x;\lambda)$ is calculated by the sum-product loopy belief propagation (LBP) [30], [57].

5 Edge-Based Image Segmentation Using Bayesian Network

The CRF naturally models the region-based image segmentation. For the edge-based image segmentation, we use a multilayer BN to model it. Bayesian Network [29], [30] is a directed acyclic graph (DAG) that consists of a set of random variables and a set of directed links between these random variables. Each variable (node) has a set of mutually exclusive states. The directed links represent the causal dependence between the random variables. The probability distribution of a node is defined by its conditional probability distribution, given the states of its parents. With a BN, the joint probability of a set of random variables can be factored into the product of a set of local conditional probabilities that are easier to compute/estimate.

We use the BN to explicitly capture the causal relationships that naturally exist among image entities such as regions, edges, vertices, and their image measurements. For example, intersections of regions naturally produce edges, while interactions of edges yield vertices. The BN provides a probabilistic framework that can systematically combine the image observations and various causal relationships so that image segmentation can be performed through a principled probabilistic inference. Besides, the BN provides a direct analogy to the human reasoning process. It can straightforwardly represent the causalities that have been extensively exploited by human. This advantage is unique for the BN compared to those undirected graphical models (e.g., CRF and MRF).

The whole structure of our BN model consists of multiple layers. Each layer will be described in the following sections.

5.1 Region, Edge, and Vertex Nodes

We build the BN model based on an oversegmented edge map that is also used to construct the superpixel CRF model in Section 4. The edge map consists of edge segments $\{e_j\}_{j=1}^m$ and vertices $\{v_t\}_{t=1}^l$, where m is the number of edges and l is the number of vertices. In this work, a vertex is the place where three or more edges intersect, i.e., a junction.

The BN model represents the relationships among the superpixel region nodes $\{y_i\}$, the edge segments $\{e_j\}$, and the vertices $\{v_t\}$. We use a synthetic edge map in Fig. 3a to explain how to construct the BN model. The basic BN model consists of three layers as shown in Fig. 3b. Specifically, the region node layer contains all of the superpixel region nodes. The edge node layer contains all the edge segments. The vertex node layer contains all vertices that are the intersections of edges.

The parents of an edge node are the two regions that intersect to form this edge. If the parents of an edge e_j have different labels, it is more likely that there is a true object boundary between them, i.e., $e_j = 1$. The relationship between the edge node e_j and its parent region nodes $pa(e_j)$ is defined by the conditional probability $P(e_j|pa(e_j))$. The conditional probability $P(e_j|pa(e_j))$ is defined as follows:

$$P(e_j = 1|pa(e_j)) = \begin{cases} 0.8, & \text{if the parent region labels are different;} \\ 0.2, & \text{otherwise.} \end{cases} \tag{9}$$

This definition basically means that the edge segment e_j has a high probability of being a true boundary when the two adjacent regions are assigned different region labels.

The edge nodes and the vertex nodes are causally linked, too. The parents of a vertex node are those edges that intersect to form this vertex. Each edge node is a binary node. Its true state represents that this edge segment belongs to the object boundary. The vertex node also assumes binary values (true or false) and it is true if the vertex is actually a corner on the object boundary.

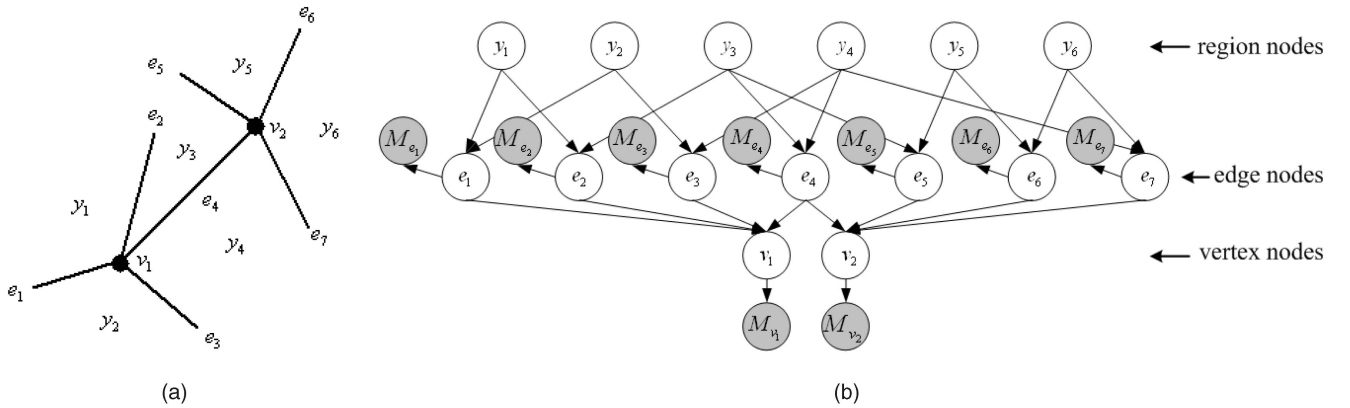

Fig. 3. A synthetic edge map and the BN that models the statistical relationships among superpixel regions, edge segments, vertices, and their measurements: (a) The synthetic edge map. There are six superpixel regions $\{y_i\}_{i=1}^{6}$, seven edge segments $\{e_j\}_{j=1}^{7}$, and two vertices $\{v_t\}_{t=1}^{2}$. (b) The corresponding basic BN structure. The shaded circles represent the measurement nodes.

Both the edge nodes and the vertex nodes have image measurements (i.e., the shaded nodes in Fig. 3b). We can use complex edge features (e.g., edgelet) or the edge probability estimated by other approaches as the edge measurements. For simplicity, we use the average gradient magnitude as the edge measurement in this work. We denote the measurement of the edge node e_j as M_{e_j}. The measurement nodes $\{M_{e_j}\}$ are continuous nodes. The conditional probability $P(M_{e_j}|e_j)$ is parameterized using Gaussian distributions defined with mean μ and variance σ^2, which are learned from the training data.

Similarly, each vertex node is also associated with an image measurement. The M_{v_t} node in Fig. 3b is the measurement of a vertex node v_t. We use the Harris corner detector [58] to calculate the measurement. Let $I(r, c)$ denote the corresponding gray-scale image. The Harris matrix A is given by

$$A = \begin{bmatrix} \left(\frac{\partial I}{\partial r}\right)^2 & \frac{\partial I}{\partial r}\frac{\partial I}{\partial c} \\ \frac{\partial I}{\partial r}\frac{\partial I}{\partial c} & \left(\frac{\partial I}{\partial c}\right)^2 \end{bmatrix}, \tag{10}$$

where r and c denote the row and column coordinates, respectively. Given the Harris matrix A, the strength of a corner is determined by a corner response function $R = det(A) - k \cdot trace(A)^2$, where k is set as the suggested value 0.04 [58].

The vertex measurement M_{v_t} is currently discretized according to the corner response R. If the corner response R is above a threshold (fixed as 1,000) and it is a local maximum, a corner is detected and the measurement node M_{v_t} becomes true. If no corner is detected at the location of the vertex v_t, the measurement node M_{v_t} becomes false. The conditional probability $P(M_{v_t}|v_t)$ quantifies the statistical relationship between the vertex node v_t and its measurement M_{v_t}. It describes the uncertainty between the state of a vertex and its measurement. This conditional probability is defined based on the empirical distribution of the measurements.

5.2 Local Smoothness Constraint

In the real world, the boundary of a natural object is locally smooth. We enforce the local smoothness constraint in our BN model by penalizing sharp corners between the intersecting edges. A sharp corner is defined as an angle between two intersecting edges that is less than $\frac{\pi}{6}$. In order to impose this constraint, the angular node $\omega_{j,s}$ is introduced to model the relationship between two adjacent edges e_j and e_s. The parent nodes (e_j and e_s) of the angular node $\omega_{j,s}$ correspond to the edge segments that intersect to form this angle. The angular node $\omega_{j,s}$ is a binary node, with its true state meaning that the local smoothness constraint is violated by these two edges. Fig. 4 illustrates how the angular nodes are added into the BN model. The measurement $M_{\omega_{j,s}}$ of an angle node is currently discretized according to a small angle. If the angle $\omega_{j,s}$ is smaller than

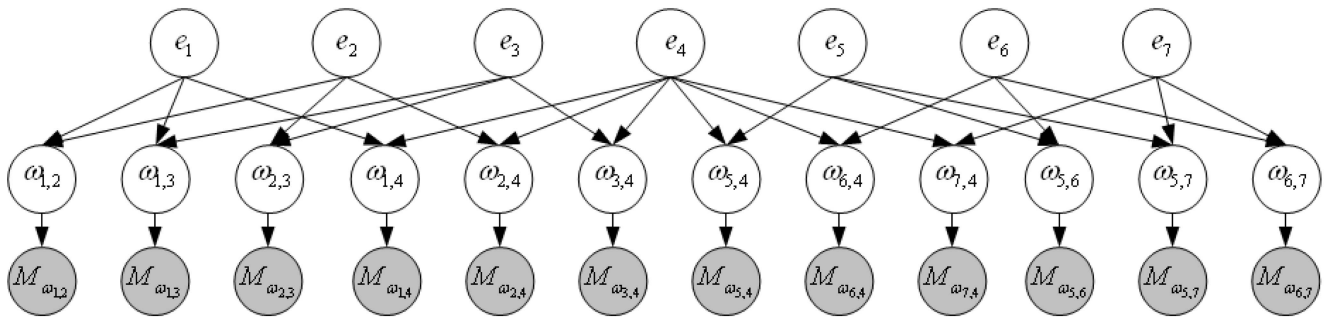

Fig. 4. The BN model with angular nodes to impose the local smoothness constraint on the edge nodes. There is an angular node corresponding to the angle between two intersecting edge segments. The parent edge nodes of an angular node correspond to those intersecting edge segments.

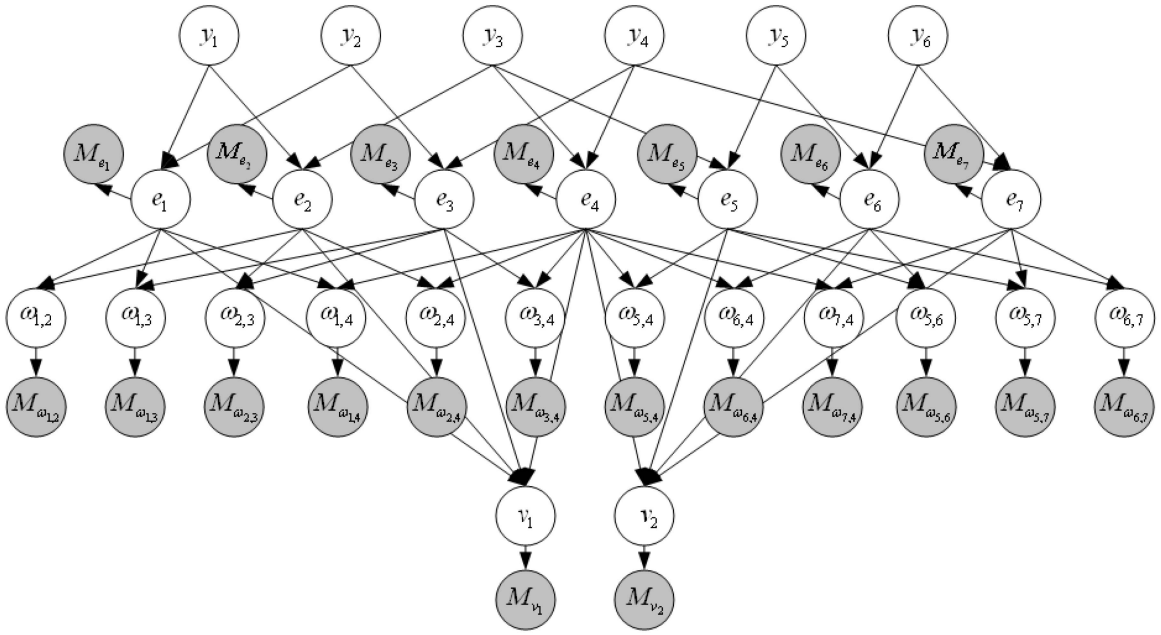

Fig. 5. The complete Bayesian Network model for the example in Fig. 3a. It consists of six region nodes $\{y_i\}_{i=1}^{6}$, seven edge nodes $\{e_j\}_{j=1}^{7}$, 12 angular nodes, two vertex nodes $\{v_t\}_{t=1}^{2}$, and the measurements of edge nodes, angular nodes, and vertex nodes.

$\frac{\pi}{6}$, the measurement node becomes 1 (true). The conditional probability table (CPT) between an angular node and its measurement can be set according to the empirical distribution of angle measurements.

To enforce the smoothness constraint, a CPT is defined to specify the relationship between the angular node $\omega_{j,s}$ and the edges e_j and e_s that intersect to form this angle, i.e.,

$$P(\omega_{j,s}=1|e_j,e_s)=\begin{cases}0.2, & \text{if both } e_j \text{ and } e_s \text{ are 1;}\\ 0.5, & \text{otherwise.}\end{cases} \tag{11}$$

This conditional probability definition effectively reduces the probability that both e_j and e_s are true boundary edges if the angle between them is too small (i.e., $\omega_{j,s}=1$). In other words, it penalizes the existence of a sharp corner in the object boundary.

5.3 Connectivity Constraint

In general, the boundary of an object should be simply connected, i.e., an edge segment should connect with at most one edge segment at its end points. This constraint is imposed by defining a CPT between the edge nodes and the related vertex node as follows:

$$P(v_t=1|pa(v_t))=\begin{cases}1, & \text{if exactly two parent edge nodes are true;}\\ 0.3, & \text{if none of the parent edge nodes is true;}\\ 0, & \text{otherwise,}\end{cases} \tag{12}$$

where $pa(v_t)$ denotes all of the parent edge nodes of the vertex node v_t.

If a corner is detected, the measurement M_{v_t} becomes true. The vertex node v_t will have a high probability of being true. In such a case, it is most likely that exactly two parent edge nodes are true (i.e., on the boundary), which corresponds to the first case of the CPT definition in (12). This implies the simple connectivity of edges at this vertex. In the second case of the CPT definition in (12), we set the entry 0.3 to account for the case when corners are detected in the background. In such a case, it is possible that none of the parent edge segments is the true object boundary. However, the conditional probability for this case shall be smaller than the case that exactly two parent edge nodes are true.

Given the CPT definition in (12), the connectivity constraint is imposed into the BN model because it favors the case that exactly two intersecting edges are the true object boundaries.

5.4 Bayesian-Network-Based Image Segmentation

The complete BN model for the synthetic example in Fig. 3a is shown in Fig. 5. Based on the BN model, the goal of image segmentation is to infer the states of the edge nodes $\{e_j\}_{j=1}^{m}$, given various measurements and constraints.

Let $\mathbf{e}$ represent all the edge nodes $\{e_j\}_{j=1}^{m}$ and $\mathbf{y}$ represent all the region nodes $\{y_i\}_{i=1}^{n}$. Similarly, ω represents all the angular nodes $\{\omega_{j,s}\}$ and $\mathbf{v}$ represents all the vertex nodes $\{v_t\}_{t=1}^{l}$. Let $\mathbf{M_e}$ represent all of the measurements for the edge nodes. $\mathbf{M_v}$ represents all of the measurements for the vertex nodes and $\mathbf{M}_\omega$ represents all the measurements for the angular nodes. Image segmentation can be performed by searching for the most probable explanation (MPE) of all hidden nodes in the BN model given various measurements, i.e.,

$$\begin{aligned}e^*,y^*,\omega^*,v^* &= \arg\max_{e,y,\omega,v} P(e,y,\omega,v|M_e,M_\omega,M_v)\\ &= \arg\max_{e,y,\omega,v} P(e,y,\omega,v,M_e,M_\omega,M_v).\end{aligned} \tag{13}$$

In the MPE results, the edge nodes with true states form the object boundary that we are looking for.

We can calculate the joint probability of all nodes in the BN model as follows:

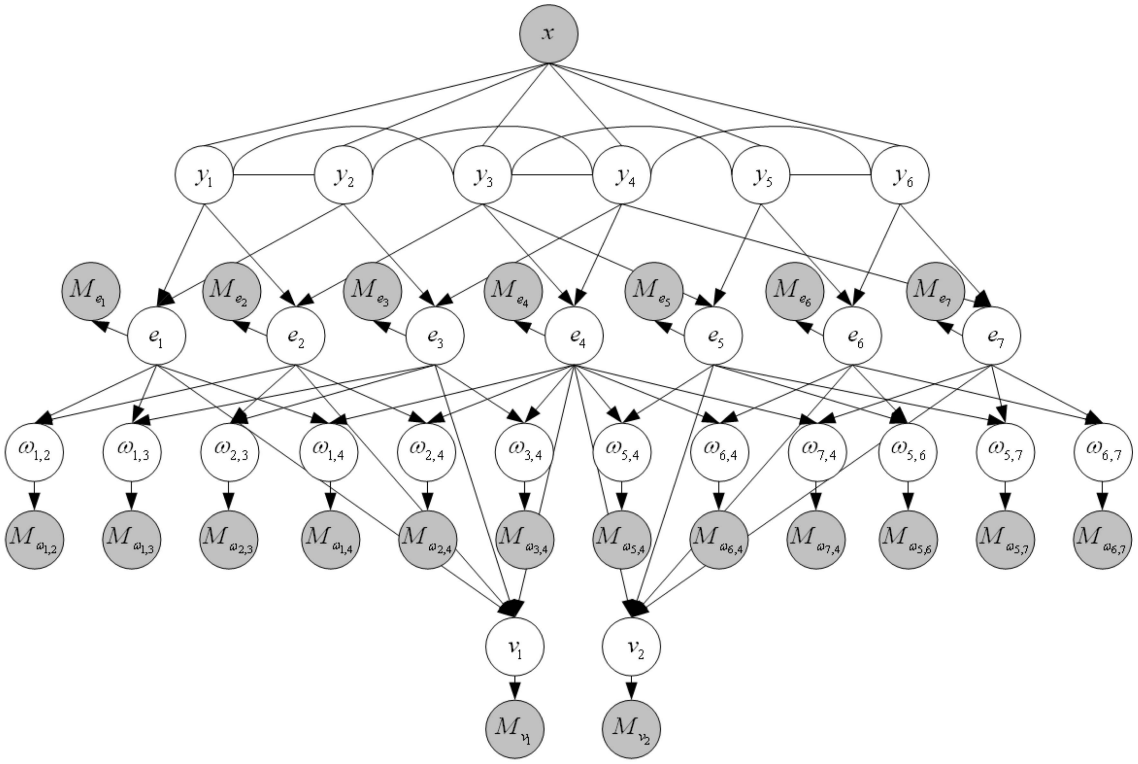

Fig. 6. The unified graphical model corresponding to the example in Fig. 3a. It combines the CRF model with the BN model through the region nodes.

$$P(e, y, \omega, v, M_e, M_\omega, M_v) = \prod_{i=1}^{n} P(y_i) \prod_{j=1}^{m} P(e_j|pa(e_j)) P(M_{e_j}|e_j) \prod_{j=1,s\in\Omega_j}^{m} P(\omega_{j,s}|e_j, e_s) P(M_{\omega_{j,s}}|\omega_{j,s}) \prod_{t=1}^{l} P(v_t|pa(v_t)) P(M_{v_t}|v_t), \quad (14)$$

where $pa(e_j)$ denotes the parent nodes of e_j. $pa(v_t)$ denotes the parent nodes of v_t. Ω_j denotes the set of edges that intersect with the edge e_j. The factorization of this joint probability is based on the conditional independence relationships among all nodes, which are implied by the constructed BN model.

Among the factored probabilities, $P(y_i)$ is the prior probability of the region nodes. Without additional prior information, it is modeled as a uniform distribution. Other terms in (14) are already defined in Sections 5.1, 5.2, and 5.3. Given all of these terms, the joint probability in (14) can be calculated. The most probable states of the edge nodes can be found using a probabilistic inference approach to find the MPE solution. Specifically, the Junction Tree method [29] is used to find the exact MPE solution in the BN model.

5.5 Parameter Estimation

Since each node in a BN is conditionally independent of other nodes that are not within its Markov blanket [29], we can locally learn the conditional probability distributions (CPDs) of each node. For example, we can learn the Gaussian distributions of the edge measurements, i.e., the likelihood model $P(M_{e_j}|e_j)$. Given the training images and their manual labeling, we classify the edge segments in the manually labeled images into two sets: the boundary edges and the nonboundary edges. We fit Gaussian distributions for the edge measurements in each set to learn $P(M_{e_j}|e_j)$. In a similar way, we also learned the likelihood models of other measurements (i.e., $P(M_{v_t}|v_t)$ and $P(M_{\omega_{j,s}}|\omega_{j,s})$). The learned conditional probabilities are then used in (14) to calculate the joint probability.

It is possible to learn the remaining CPDs in a similar way. However, the training on a specific data set tends to skew the BN model only for that specific set of training data. Such a trained BN model cannot generalize well to other unseen data. As a result, we opt for a soft BN parameterization instead of a hard BN parameterization. We empirically set the fixed values for some conditional probability parameters (as shown in previous equations) due to several considerations. First, we can directly define those CPDs according to the semantic meaning of their conditional probabilities. Second, some previous work [59] shows the performance of BNs for diagnosis is not very sensitive to the accurate parameter setting. Third, we have changed the CPDs (e.g., $P(e_j|pa(e_j))$) for all edge nodes within a range of ± 10 to 20 percent relative to the preset values. The segmentation results did not change very much, which agrees with the observations in [59]. In Section 7.1, we will show a set of experiments to demonstrate this phenomenon. Fourth, we applied the model using this parameterization strategy on different data sets and found it generally performed well.

6 A Unified Graphical Model Combining the CRF Model with the BN Model

The complete segmentation model unifies the CRF part and the BN part through the causal relationships between the region nodes $\mathbf{y}$ and the edge nodes $\mathbf{e}$, as shown in Fig. 6.

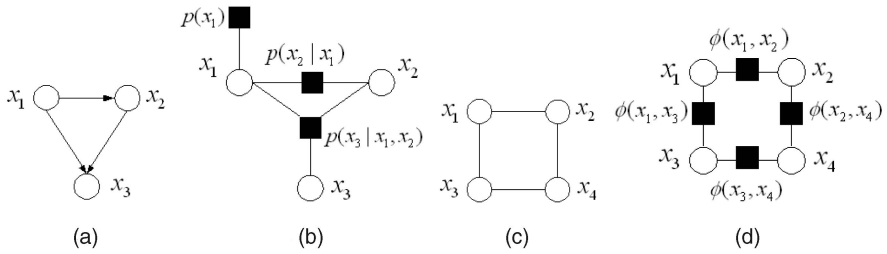

Fig. 7. Examples of different types of graphical models and their corresponding factor graph representations: (a) Bayesian Network and (b) its corresponding factor graph representation; (c) Markov Random Field and (d) its corresponding factor graph representation.

The CRF part collects region-based information to infer the states of the region nodes. The BN part collects edge-based information and the imposed local constraints to infer the states of the edge segments. The unified model therefore integrates two kinds of complementary image segmentation principles in a single framework. More importantly, it exploits both the spatial homogeneity of the region labels and the hierarchical causal relationships among regions, edges, and vertices for image segmentation.

The unified graphical model in Fig. 6 consists of both directed links and undirected links. To perform a consistent inference, it is necessary to convert the unified model into a Factor Graph (FG) representation [60], [61] since it is very difficult to directly perform inference in such a graphical 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 factored 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 factored function. Following the convention, each variable node will be represented as a circle and each factor node will be represented as a filled square in the factor graph.

Since both the undirected graphical model (e.g., MRF) and the directed graphical model (e.g., BN) represent the factored joint probability distribution of a set of variables, they can be easily converted into a factor graph representation [60], [61], [31]. Fig. 7a is a BN that represents a joint probability distribution $P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$. Based on this factorization, a factor graph can be constructed to model the same distribution, as shown in Fig. 7b. The variables in the FG correspond to those variables in the BN. Factor nodes are added to correspond to the factored probabilities in the joint probability. Edges are added to link a factor node and a variable node if and only if the factor is a function of this variable. We can convert an undirected graphical model into a factor graph in a similar way. Fig. 7c is a simple MRF that represents a joint distribution $P(x_1, x_2, x_3, x_4) = \phi(x_1, x_2)\phi(x_1, x_3)\phi(x_3, x_4)\phi(x_2, x_4)$, where the normalization constant can be merged into one factor such as $\phi(x_1, x_2)$. Based on this factorization, a factor graph is constructed in a similar way to model the same distribution, as shown in Fig. 7d. Each factor node corresponds to the factored potential function.

Based on the graphical structure of the unified model in Fig. 6, we can factorize the joint probability distribution of all the variables according to the global Markov property in the graphical model (cf. [28, Chapter 3]). According to the graphical structure, $\mathbf{x}$ are conditionally independent of other random variables given $\mathbf{y}$. Hence, we have

$$
\begin{aligned}
&P(\mathbf{y}, \mathbf{e}, \omega, \mathbf{v}, \mathbf{M_e}, \mathbf{M}_\omega, \mathbf{M_v}, \mathbf{x}) \\
&= P(\mathbf{e}, \omega, \mathbf{v}, \mathbf{M_e}, \mathbf{M}_\omega, \mathbf{M_v}|\mathbf{y}, \mathbf{x})P(\mathbf{y}, \mathbf{x}) \\
&= P(\mathbf{e}, \omega, \mathbf{v}, \mathbf{M_e}, \mathbf{M}_\omega, \mathbf{M_v}|\mathbf{y})P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \\
&= P(\mathbf{M_e}, \mathbf{M}_\omega, \mathbf{M_v}|\mathbf{e}, \omega, \mathbf{v}, \mathbf{y})P(\mathbf{e}, \omega, \mathbf{v}|\mathbf{y})P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \\
&= P(\mathbf{M_e}|\mathbf{e})P(\mathbf{M}_\omega|\omega)P(\mathbf{M_v}|\mathbf{v})P(\omega, \mathbf{v}|\mathbf{e}, \mathbf{y})P(\mathbf{e}|\mathbf{y})P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \\
&= P(\mathbf{M_e}|\mathbf{e})P(\mathbf{M}_\omega|\omega)P(\mathbf{M_v}|\mathbf{v})P(\omega|\mathbf{e})P(\mathbf{v}|\mathbf{e})P(\mathbf{e}|\mathbf{y}) \\
&\quad P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \\
&= \frac{1}{\mathbb{Z}} \prod_{j=1}^{m} P\big(M_{e_j}|e_j\big) \prod_{j=1, s\in\Omega_j}^{m} P\big(M_{\omega_{j,s}}|\omega_{j,s}\big) \prod_{t=1}^{l} P\big(M_{v_t}|v_t\big) \\
&\qquad \prod_{j=1, s\in\Omega_j}^{m} P(\omega_{j,s}|e_j, e_s) \prod_{t=1}^{l} P(v_t|pa(v_t)) \prod_{j=1}^{m} P(e_j|pa(e_j)) \\
&\qquad \prod_{i\in V} \phi(y_i, x_i) \prod_{i\in V} \prod_{j\in\mathcal{N}_i} \exp(y_i y_j \lambda^T g_{ij}(x)),
\end{aligned}
\tag{15}
$$

where $\mathbb{Z}$ is the normalization constant. Note that $P(\mathbf{x})$ is a constant since $\mathbf{x}$ are observed. It can therefore be merged into the normalization constant $\mathbb{Z}$. Among these factored functions, $P(M_{e_j}|e_j)$, $P(M_{\omega_{j,s}}|\omega_{j,s})$ and $P(M_{v_t}|v_t)$ are the likelihood models of the measurements of edges, angles, and vertices, respectively. $P(\omega_{j,s}|e_j, e_s)$, $P(v_t|pa(v_t))$, and $P(e_j|pa(e_j))$ are the conditional probabilities of angular nodes, vertex nodes, and edge nodes, respectively. All of these conditional probabilities are defined in the BN part of the unified model (Section 5). The remaining factored functions $\phi(y_i, x_i)$ and $\exp(y_i y_j \lambda^T g_{ij}(x))$ are the potential functions defined in the CRF part of the unified model (Section 4). Based on the factored joint probability distribution in (15), we convert the unified model into a factor graph representation in order to perform joint inference. The converted FG is shown in Fig. 8, where each factor node corresponds to one factored function in the above equation.

Given the factor graph representation, there are different principled ways to perform probabilistic inference. First, the sum-product algorithm can be used to efficiently calculate various marginal probabilities for either a single variable or a subset of variables [60], [31]. The sum-product algorithm operates via a local "message passing" scheme. It can perform exact inference of the marginal probabilities for singly connected graphs. Let $\mathbf{R}$ denotes all the variables in the above factor graph and $r_i \in \mathbf{R}$ is one of the variable. We

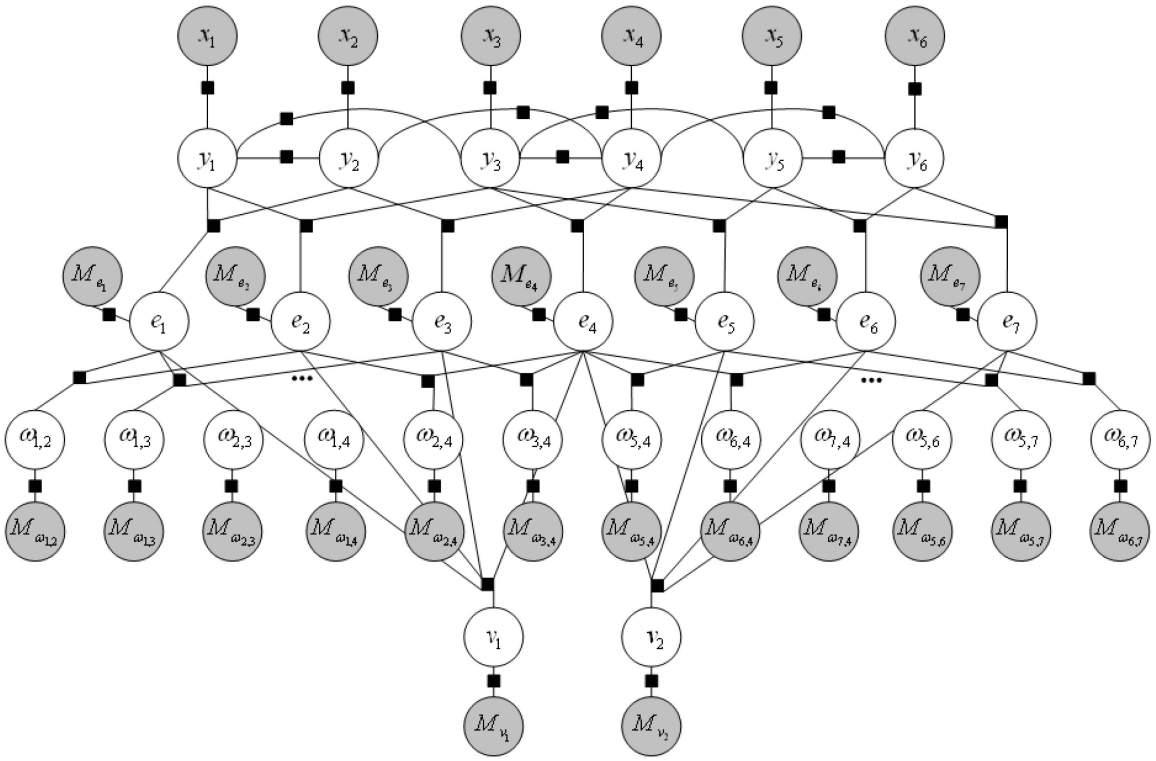

Fig. 8. The factor graph representation of the unified graphical model in Fig. 6. The circles represent the variables, while the filled squares represent the factors. The shaded circles are observed variables. For clarity, not all factors are drawn. The symbol "..." represents the undrawn factors.

can use the sum-product algorithm to calculate the marginal probability of r_i, i.e.,

$$P(r_i) = \sum_{\mathbf{R}\backslash r_i} P(\mathbf{R}),$$

where the summation is over all variables excluding r_i. $P(\mathbf{R})$ is the joint probability in (15). Given the marginal probability of each variable, the optimal state of the variable can be found by the MPM criterion [56], i.e.,

$$r_i^* = \arg\max_{r_i} P(r_i).$$

Second, the max-product algorithm [31] can be used to find a setting of all of the variables that corresponds to the largest joint probability, i.e.,

$$\mathbf{R}^* = \arg\max_{\mathbf{R}} P(\mathbf{R}).$$

This optimal solution has the same meaning as the MPE solution mentioned in the Bayesian Network inference (Section 5.4). The max-product algorithm works identically to the sum-product algorithm except that the summation in the sum-product algorithm is replaced by the maximization when we calculate the messages.

Third, besides the max-product algorithm, there are other algorithms that can also find the MPE solution given the evidence. The stochastic local search (SLS) [62] is one such algorithm. In [63], Hutter et al. improve Park's algorithm [62] to achieve a more efficient algorithm for MPE solving. They also extend this algorithm and provide a public available software to deal with various types of graphical models, including the factor graph. Given the FG model in Fig. 8, we use the inference package provided by Hutter et al. to perform MPE inference in the factor graph, i.e.,

$$y^*, e^*, \omega^*, v^* = \arg\max_{y,e,\omega,v} P(\mathbf{y}, \mathbf{e}, \omega, \mathbf{v}, \mathbf{M_e}, \mathbf{M}_\omega, \mathbf{M_v}, \mathbf{x}), \quad (16)$$

where the joint probability is calculated by (15). In the MPE solution, the region nodes with the foreground labels form the final segmentation.

7 Experiments

7.1 Figure/Ground Image Segmentation

We first tested the proposed model for figure/ground image segmentation using the Weizmann horse data set [32]. This data set includes the side views of many horses that have different appearances and poses, which makes it challenging to segment these images. On the other hand, several related works [32], [23], [64], [65], [20], [16], [66], [11] also did experiments on this data set. We can compare our results with these state-of-the-art works.

Our approach requires a learning process to train the model. For this purpose, we used 60 horse images as the training data. The test images include 120 images from the Weizmann horse data set. 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, many of which have never been seen in the training set. Due to these reasons, the segmentation is a challenging problem.

We have performed the experiments using both color images and gray-scale images of the same training set and the testing set. It demonstrates that our approach is flexible

Fig. 9. Examples of the color image segmentation results arranged in two groups of two rows. In each group, the first row includes the color horse images. The second row includes the segmentation masks produced by the proposed approach.

enough to segment different types of images. Different features have been used in the region-based CRF part for segmenting the color images and the gray-scale images. For the color images, we use the average CIELAB color and their standard deviations as the local features x_i for each superpixel region. In this case, the length of the feature vector is 6. 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 the gray-scale images, we use the average intensity and 12 Gabor textures as the features for each superpixel region. The Gabor textures are calculated by filtering the gray-scale image with a set of Gabor filter banks. The average magnitude of the filtered image in each superpixel region is used as the Gabor feature. We use the Gabor filter banks with three scales and four orientations. In this case, the length of the feature vector x_i is 13. The three-layer perceptron has a structure with 13 nodes in the input layer, 25 nodes in the hidden layer, and 1 node in the output layer. We are using fewer hidden nodes because the number of nodes in the input layer is increased but the total number of training data remains the same.

All of the training images and the test images are first oversegmented using the Edgeflow-based anisotropic diffusion method. Given the training images and their ground truth labeling, we automatically train the unified graphical model using the process described in the Sections 4.2 and 5.5.

After learning the model, we perform image segmentation on the test images using the inference process described in Section 6. Fig. 9 shows some examples of the color horse images and their segmentation masks. Fig. 10 shows examples of the gray-scale horse images and their segmentation masks. We achieved encouraging results on these images. Most small errors happen on the horse's feet, where the appearances of these parts are different from the horse's body. Another kind of error is caused by the clutter. When the background (e.g., the shadow) has a similar appearance as the foreground, the proposed model may not be able to completely separate them.

We first qualitatively compare our segmentation results with some results produced by other state-of-the-art approaches. Cour and Shi [64] segment an image by finding the optimal combination of the superpixel regions in an oversegmentation produced by the Normalized Cuts [3]. The optimal segmentation in their approach shall have a similar shape as the shape template that is generated from the manually labeled training data. Borenstein et al. [32], [23] combine the top-down and bottom-up process to perform image segmentation. Levin and Weiss [65] propose a learning-based approach for image segmentation. They perform image segmentation based on matching patch templates with the images. Winn and Jojic [20] learn the object class model from the unlabeled images. Their approach combines the bottom-up cues of color and edge with top-down cues of shape and pose. Zhu et al. [66] propose an unsupervised structure learning method to learn the hierarchical compositional model for deformable objects and apply it to segment articulated objects. All of these works have been tested on (a subset of) the Weizmann horse data set. Fig. 11 shows several example segmentation results of these approaches on color images. Our results can compete with these results according to the visual inspection.

In order to quantitatively evaluate our segmentation results and compare with the aforementioned approaches, we calculate the average percentage of correctly labeled pixels (i.e., segmentation consistency [67]) in all test images. The quantitative results are summarized in Table 1. In this table, we also list the segmentation consistency achieved by other related works.

From the quantitative results in Table 1, we conclude that our results are comparable to (or better than) the results produced by other state-of-the-art approaches. Note that we

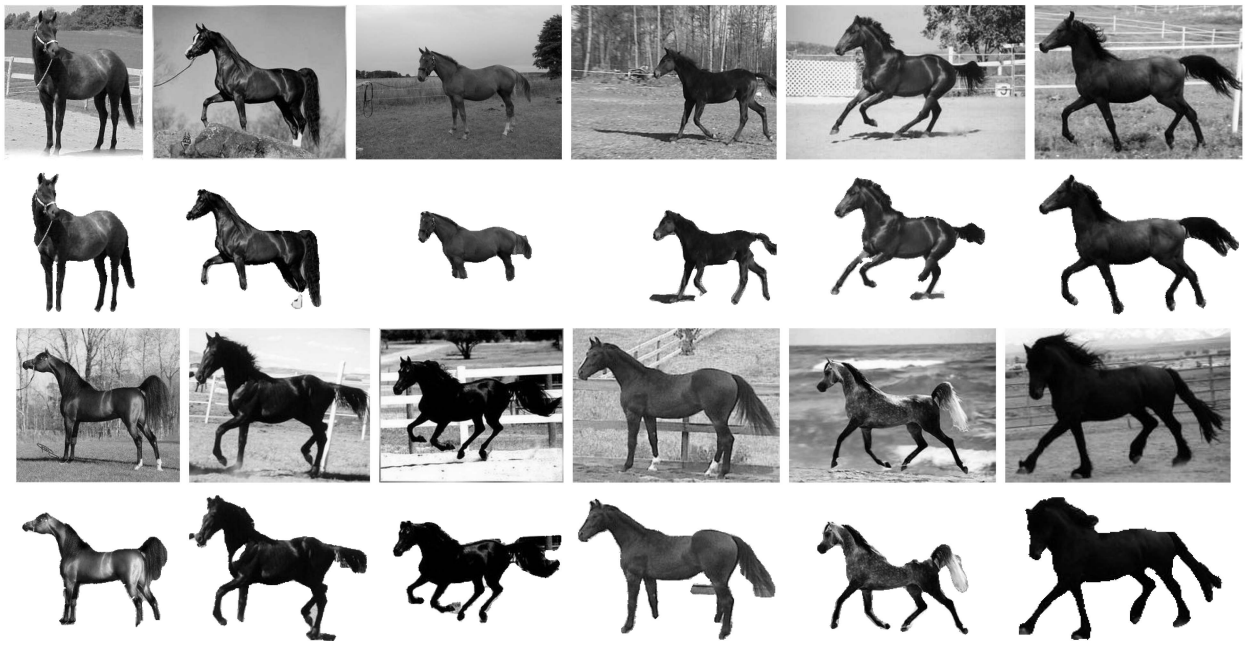

Fig. 10. Examples of the gray-scale image segmentation results arranged in two groups of two rows. In each group, the first row includes the gray-scale test images. The second row includes the segmentation masks produced by the proposed approach.

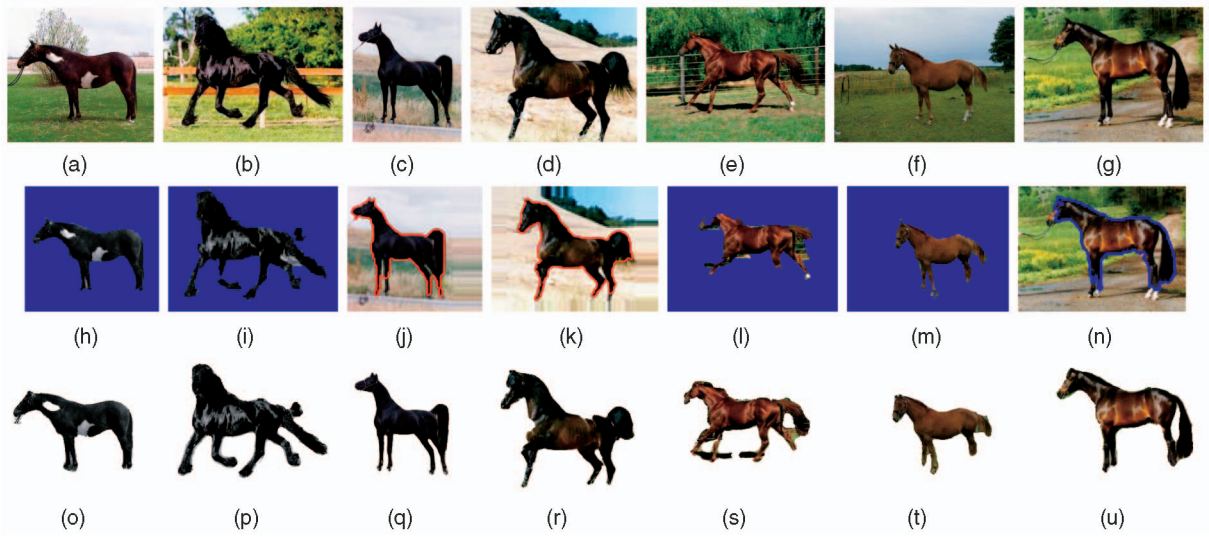

Fig. 11. Qualitative comparison of our segmentation results with the results produced by other state-of-the-art approaches [23], [65], [64], [20]. (a)-(g) The test images. (h)-(n) The screen copies of some results from the related works: (h) and (i) are the results from [23]. (j) and (k) are the results from [65]. The segmentation is superimposed as red contours. (l) and (m) are the results from [64]. (n) is the result from [20]. The segmentation is superimposed as blue contours. (o)-(u) The corresponding results produced by our approach. This figure should be viewed in color.

only use very simple image features (e.g., the color) for segmentation and have not performed any feature selection. Besides, we have not utilized the additional object shape information as some works have done [64], [23]. We notice that the work [65] has performed the feature selection from a pool of 2,000 features, which may be crucial to increasing its performance. In addition, they only show segmentation results on eight testing images in their paper. They have not mentioned how many images they have actually used for testing. Therefore, the performance of their approach on a relatively large data set is unknown. It is also difficult to directly compare our results with the results reported in [32], [16] because they did not give the segmentation consistency measurement.

In Table 1, we also list the performance using a CRF model alone and using a BN model alone. The CRF model is exactly the same as what is described in Section 4. The BN model is basically the same as what is described in Section 5. We further incorporate the CIELAB color as the measurements of the region nodes. The likelihood of the region measurements given the region label is simply modeled as Mixture of Gaussians, which are learned from the training data. Compared with the performance of the unified model, it is apparent that the unified model combining both the CRF part and the BN part performs much better than either using the CRF model alone or using the BN model alone. These results demonstrate the usefulness of both parts in our unified model.

TABLE 1
The Quantitative Comparison of Our Approach with Several Related Works for Segmenting the Weizmann Horse Images

method	image type	train	test	segmentation consistency
Cour et al. [64]	color	20	308	94.2%
Levin et al. [65]	color	N/A	N/A	95.0%
Winn et al. [20]	color	20	200	93.1%
Zhu et al. [66]	color	12	316	93.3%
Ren et al. [11]	color	172	172	91.0%
Borenstein et al. [23]	grey scale	64	328	93.0%
Winn et al. [20]	grey scale	20	200	93.0%
our unified model	grey scale	60	120	94.0%
our unified model	color	60	120	95.4%
our CRF model alone	color	60	120	92.5%
our BN model alone	color	60	120	93.7%

The average percentage of correctly labeled pixels (i.e., segmentation consistency) is used as the quantitative measurement.

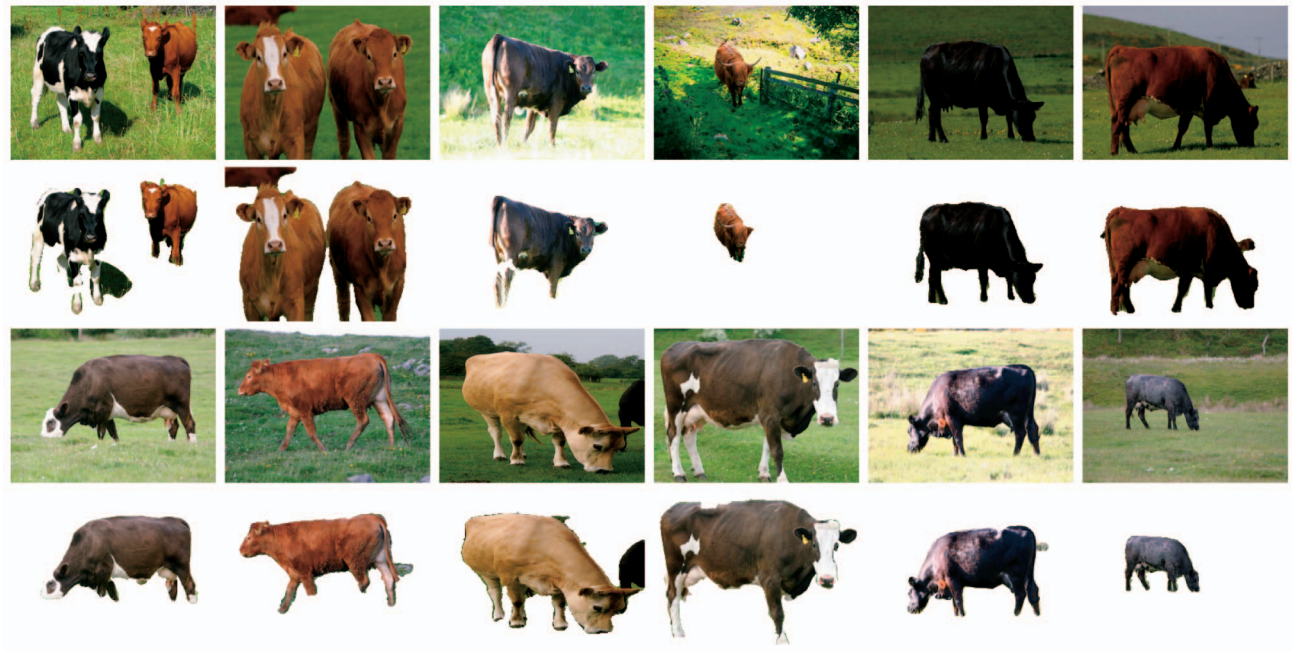

Fig. 12. Examples of the segmentation results of the VOC2006 cow images arranged in two groups of two rows. In each group, the first row includes the color test images. The second row includes the segmentation masks produced by the proposed approach.

For the figure/ground image segmentation, we have also performed the experiments on a set of cow images from the VOC2006 database [34]. This database is primarily used for object categorization. In this work, we use it to test our image segmentation approach. 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 cow images (57 images) for training our unified model and use the rest half set of images (51 images) for testing. Some examples of the image segmentation results are shown in Fig. 12. We have achieved reasonable segmentation results. Although those cows have different appearances and sizes and there might be multiple cows in the image, our approach successfully segments them out. Besides the qualitative evaluation of these results, we manually segment the test images and use the manual segmentation as the ground truth to calculate the segmentation consistency. We achieve a good segmentation consistency of 96 percent on these cow images. We also use the CRF part in our unified model to segment these images. The CRF model alone achieved a segmentation consistency of 93.9 percent, which is apparently inferior to the unified model. This also demonstrates the benefits of unifying two parts for improved performance. Specifically, we observed that the CRF model alone tended to oversmooth the labeling without the help of the BN part.

Besides providing the segmentation consistency on the two sets of figure/ground segmentation experiments, we also summarize the quantitative results using the two-class confusion matrices in Table 2. Each row is normalized w.r.t. the total number of pixels in the foreground or in the background, respectively. Therefore, the summation of each row will be equal to 1. These confusion matrices show that most of the foreground and background are correctly labeled.

In Section 5.5, we have mentioned that the performance of the model is not very sensitive to the accurate parameter setting in the Bayesian Network part. We perform a set of

TABLE 2
Two-Class Confusion Matrices of the Figure/Ground Segmentation on (a) the Weizmann Data Set and (b) the VOC2006 Cow Data Set

ground truth\segmentation	foreground	background
foreground	**91.8%**	8.2%
background	3.4%	**96.6%**

(a)

ground truth\segmentation	foreground	background
foreground	**93.1%**	6.9%
background	2.9%	**97.1%**

(b)

TABLE 3
The Segmentation Consistencies When the Conditional Probability Table $P(e_{ij}|y_i, y_j)$ in the Model Changes

$P(e_{ij} = 1 \mid y_i \neq y_j)$	0.7	0.75	0.8	0.85	0.9	0.95
Weizmann dataset	94.5%	94.9%	95.4%	95.5%	94.7%	94.4%
VOC2006 dataset	96.4%	96.4%	96%	96%	95.5%	95.3%

Note that $P(e_{ij} = 1|y_i \neq y_j)$ shall be larger than 0.5 because there is more likely a boundary ($e_{ij} = 1$) when the adjacent regions have different labels.

TABLE 4
The Segmentation Consistency on the VOC2006 Data Set When Different Methods Are Used for the Initial Oversegmentation

oversegmentation method	anisotropic diffusion [53]	Normalized Cuts [3]
segmentation consistency	96%	95.7%

experiments to validate this. We change the conditional probability $P(e|pa(e))$ for all edge nodes and redo the segmentation of color images in the Weizmann data set and VOC2006 cow data set. Specifically, we change $P(e_{ij} = 1|y_i \neq y_j)$ (and the related CPT entries) to six different values but retain all other configurations. The segmentation consistencies of this set of experiments are summarized in Table 3. We observe that the overall performance only changes about one percent, which shows the performance is not very sensitive to the accurate parameter setting in the BN part. However, we also notice that extremely inappropriate parameter setting may decrease the performance.

In addition, although we use the anisotropic segmentation software to produce the initial oversegmentation, we can also use other approaches to produce the oversegmentation. For example, we use the public available Normalized Cuts software to oversegment the image into 50 segments. We then redid the segmentation on the VOC2006 cow data set and found that the segmentation consistency just slightly changed, as shown in Table 4. It demonstrates that our approach does not depend on a specific method for the initial oversegmentation.

Finally, since an oversegmentation is required to produce the input edge map for constructing the proposed model, we did a set of experiments to study the influence of the initial oversegmentation on the overall segmentation performance. We use Normalized Cuts software to produce the oversegmentation of the VOC2006 cow images with 50, 60, 80, 100, and 120 number of oversegmentations. Then, we use the same model with the same parameters to segment these images. The quantitative results of these experiments are summarized in Fig. 13. We observed that the initial oversegmentation with different numbers only has marginal influence on the overall segmentation performance. However, if the initial oversegmentation is too coarse, there will be segments crossing the object boundary and leading to incorrect segmentation. In practice, we observed that the anisotropic segmentation software can produce oversegmentation where the edges align with the true object boundary well. On the other hand, if the oversegmentation is too fine grained, the number of neighbors of a superpixel will significantly increase. This might influence the spatial interaction between adjacent superpixels and slightly change the segmentation performance.

7.2 Multiclass Segmentation

To further test the capability of the proposed model, we apply it to a difficult multiclass segmentation problem on the Microsoft data set (MSRC2) [33]. This data set includes 591 images with 21 object classes, 1 nonsense class, and 2 not-considered classes. There are significant overlaps between the appearances of different object classes (e.g., building and road). In addition, the within-class appearances also have significant variations. These reasons make the multiclass

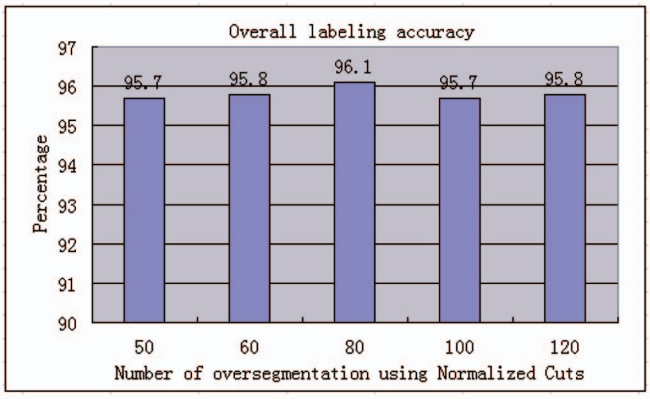

Fig. 13. The influence of the initial oversegmentation on the overall segmentation performance on the VOC2006 cow data set. Normalized Cuts software is used to produce the initial oversegmentation with different number of segments.

Fig. 14. Examples of the segmentation results of MSRC2 images arranged in three rows. (a) The color test images. (b) The multiclass labeling results produced by the proposed approach. Each color represents a different object class. (c) The ground truth labeling, where the black color indicates the nonsense class. This figure should be viewed in color.

segmentation very challenging. Even the state-of-the-art approach [68] can only achieve a 75.1 percent overall pixelwise labeling accuracy.

Although our unified model is designed as a figure/ground segmentation model, it can still be applied to multiclass segmentation. We first train 21 figure/ground segmentation models for all the object classes. We roughly divided the whole data set into two halves and use one half for training (296 images) and the other half for testing (295 images). Since some object classes rarely exist in the whole data set, we use more positive samples to train the corresponding model but ensure no overlap between the training and testing images. After the model training, we sequentially apply these models to each testing image to achieve multiclass segmentation.

In addition, since there is significant between-class overlap of object appearances and within-class variation of object appearances, we use more local features for the multiclass segmentation. Specifically, we use the color features together with 38 features calculated from Maximum Response (MR) filter sets [69]. Fig. 14 shows a few examples of the multiclass segmentation results, where different color corresponds to different object classes. We successfully segmented the multiple object classes in these images.

To quantitatively evaluate the performance, we calculate the confusion matrix of the multiclass segmentation results and the overall labeling accuracy. We achieve an overall pixelwise labeling accuracy of 75.4 percent. There are some state-of-the-art approaches that have also been tested on this data set [33], [68], [70]. We summarize the overall labeling accuracy of these approaches in Table 5. Our overall performance is slightly better than these approaches. We notice that [68] uses some implicit shape information and [70] exploits a feature selection process. These additional processes are important to achieve their performance. In contrast, we have not exploited these optional processes yet.

TABLE 5
The Quantitative Comparison of Our Approach with Several Related Works for Segmenting the MSRC2 Data Set

algorithm	overall accuracy
TextonBoost [33]	72.2%
Yang et al. [68]	75.1%
Auto-Context [70]	74.5%
Our approach	75.4%

The average percentage of correctly labeled pixels is used as the quantitative measurement.

The confusion matrix of our multiclass segmentation results is shown in Fig. 15. Each number is the percentage of labeling normalized w.r.t. the total pixels in each class. We summarize the diagonal elements of the confusion matrix in Table 6 and compare them with those from the state-of-the-art approach [68]. The bold numbers highlight those object classes where our approach performs better. Compared to [68], our approach achieves better performance on 11 classes and ties on 1 class (i.e., water). We notice that our approach performs better in some difficult classes, such as chair, sign, etc. This may be due to the fact that our model is (potentially) capable of capturing very complex relationships between multiple image entities.

7.3 Computational Time

We implement the whole model using Matlab software. The segmentation speed mainly depends on the complexity of the constructed graphical model. The constructed factor graph usually consists of 700 to 1,500 nodes. It may take several seconds to half a minute to segment an image using the efficient factor graph inference [63] in a Pentium M 1.7 GHz laptop. We summarize the typical segmentation time required by other related works in Table 7. Compared to these works, our approach is relatively more efficient on segmenting a normal size image. This can be attributed to using superpixels as the basic units in an image and the fast MPE inference using factor graph.

8 Conclusions

To summarize, we present a new image segmentation framework based on a unified probabilistic graphical model. The proposed unified model can systematically capture the complex and heterogenous relationships among different image entities and combine the captured relationships with various image measurements to perform effective image segmentation. An image is first oversegmented to produce an edge map, from which a superpixel-based CRF and a multilayer BN are automatically constructed. The superpixel CRF model performs region-based image segmentation based on the local features and the conditioned Markovian

	building	grass	tree	cow	sheep	sky	plane	water	face	car	bike	flower	sign	bird	book	chair	road	cat	dog	body	boat
building	77	1	4	0	0	4	1	1	0	2	2	0	0	0	1	0	2	0	0	2	2
grass	1	93	4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
tree	8	9	70	0	0	6	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0
cow	17	17	5	58	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0
sheep	22	10	3	0	64	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sky	4	0	0	0	0	92	1	2	0	0	0	0	1	0	0	0	0	0	0	0	0
plane	15	7	10	0	0	10	57	0	0	0	0	0	0	0	0	0	2	0	0	0	0
water	10	0	11	0	0	4	0	70	0	0	0	0	0	3	0	0	0	0	0	0	0
face	16	0	2	0	0	0	0	0	61	0	0	1	0	0	8	0	0	0	0	12	0
car	16	0	4	0	0	1	0	0	0	69	1	1	0	0	1	0	7	0	0	1	0
bike	25	1	2	0	0	0	0	0	0	1	67	0	0	0	0	0	4	0	0	0	0
flower	9	4	9	0	0	0	0	0	0	1	0	74	0	0	3	0	0	0	0	0	0
sign	22	0	2	0	0	1	0	0	0	1	1	3	70	0	0	0	0	0	0	0	0
bird	16	7	9	0	0	5	0	10	0	0	0	0	0	47	0	0	6	0	0	0	0
book	10	1	4	0	0	0	0	0	0	0	0	2	0	0	80	0	0	0	0	3	0
chair	25	7	6	0	0	0	0	1	0	0	0	3	0	0	1	53	4	0	0	0	0
road	15	0	1	0	0	1	0	0	0	4	3	0	0	0	0	4	73	0	0	0	0
cat	27	0	2	0	0	4	0	0	0	0	0	1	0	0	6	0	7	53	0	0	0
dog	24	4	3	0	0	1	0	0	0	0	0	0	0	0	0	0	11	0	56	0	0
body	16	4	3	0	0	1	0	0	22	2	0	3	0	0	1	0	1	0	1	47	0
boat	11	0	7	0	0	3	0	34	0	0	0	1	0	0	2	0	0	0	0	0	40

Fig. 15. The confusion matrix of our multiclass segmentation results on the MSRC2 data set. The rows correspond to the ground truth classes, while the columns correspond to the labeled classes by the proposed approach. The overall pixelwise labeling accuracy is 75.4 percent.

TABLE 6
The Comparison of the Diagonal Elements in the Confusion Matrices between Our Approach and the State-of-the-Art Approach [68]

algorithm	building	grass	tree	cow	sheep	sky	plane	water	face	car	bike
Yang et al. [68]	63.1	97.9	89.5	65.7	54.1	86.2	62.7	70.9	83.2	70.5	79.6
ours	**76.7**	93.3	69.8	57.8	**64.2**	**91.9**	56.8	70.5	61.0	68.7	67.1

algorithm	flower	sign	bird	book	chair	road	cat	dog	body	boat	
Yang et al. [68]	71.3	37.9	23.2	87.9	23.2	88.2	33.1	34.1	43.2	32.4	
ours	**74.2**	**70.4**	**46.7**	80.0	**52.7**	72.6	**53.3**	**56.6**	**47.2**	**39.8**	

The bold numbers indicate where our approach performs better than the compared approach.

property. The BN model performs edge-based image segmentation based on the measurements of edges, vertices, angles, and local constraints such as the smoothness and connectivity constraints. The CRF model and the BN model are then integrated through factor graph theories to produce a unified graphical model. Image segmentation under the unified graphical model is solved through an efficient MPE inference given various image measurements.

The unified graphical model systematically combines the CRF model with the BN model. These models represent different types of graphical models. The proposed unified model can therefore flexibly model both causal and noncausal relationships among different entities in the image segmentation problem. By using the unified graphical model, both region-based information and edge-based information are also seamlessly integrated into the segmentation process. The proposed approach represents a new direction for developing image segmentation methods. The experimental results on the Weizmann horse data set, the VOC2006 cow data set, and the MSRC2 mutliclass segmentation data set show that our approach achieves the segmentation performance that can rival the competing state-of-the-art approaches. It demonstrates the promising capability of the proposed unified framework.

In this work, we use a combination of supervised parameter learning and manual parameter setting for the model parameterization. Although this method works generally well in our experiments, in the long run, it may be more desirable to directly perform joint parameter learning in the unified model, i.e., to simultaneously learn the BN and CRF parameters automatically from the training data. This is not a trivial task and requires further theoretical derivations and in-depth study of the unified graphical model. We will study this issue as part of future work.

Finally, we want to point out that the application of the unified graphical model is not limited to image or video segmentation. It can find applications in many different computer vision problems including object tracking, object recognition, activity modeling and recognition, etc.

TABLE 7
The Typical Time Required for Segmenting a Normal Size Image

algorithm	normal image size	segmentation time
TextonBoost [33]	320×213	3 minutes
Yang et al. [68]	320×213	less than 1 minute
Auto-Context [70]	300×200	30s to 70s
Our approach	320×213	less than 30s

ACKNOWLEDGMENTS

This project is supported in part by a grant from the US National Science Foundation under award number 0241182. The authors also want to thank the anonymous reviewers and the editors who gave them valuable comments on this work.

REFERENCES

[1] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 603-619, May 2002.
[2] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision,* vol. 1, pp. 321-331, 1988.
[3] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 888-905, Aug. 2000.
[4] T. Chan and L. Vese, "Active Contours without Edges," *IEEE Trans. Image Processing,* vol. 10, no. 2, pp. 266-277, Feb. 2001.
[5] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 6, pp. 721-741, Nov. 1984.
[6] C. Bouman and B. Liu, "Multiple Resolution Segmentation of Textured Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 2, pp. 99-113, Feb. 1991.
[7] D. Melas and S. Wilson, "Double Markov Random Fields and Bayesian Image Segmentation," *IEEE Trans. Signal Processing,* vol. 50, no. 2, pp. 357-365, Feb. 2002.
[8] W. Pieczynski and A. Tebbache, "Pairwise Markov Random Fields and Segmentation of Textured Images," *Machine Graphics and Vision,* vol. 9, pp. 705-718, 2000.
[9] C. D'Elia, G. Poggi, and G. Scarpa, "A Tree-Structured Markov Random Field Model for Bayesian Image Segmentation," *IEEE Trans. Image Processing,* vol. 12, no. 10 pp. 1259-1273, Oct. 2003.
[10] X. He, R. Zemel, and M. Carreira Perpinan, "Multiscale Conditional Random Fields for Image Labeling," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 695-702, 2004.
[11] X. Ren, C.C. Fowlkes, and J. Malik, "Cue Integration in Figure/Ground Labeling," *Advances in Neural Information Processing Systems,* pp. 1121-1128, MIT Press, 2005.
[12] P. Awasthi, A. Gagrani, and B. Ravindran, "Image Modeling Using Tree Structured Conditional Random Fields," *Proc. Int'l Joint Conf. Artificial Intelligence,* pp. 2054-2059, 2007.
[13] X. Feng, C. Williams, and S. Felderhof, "Combining Belief Networks and Neural Networks for Scene Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 4, pp. 467-483, Apr. 2002.
[14] E.N. Mortensen and J. Jia, "Real-Time Semi-Automatic Segmentation Using a Bayesian Network," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1007-1014, 2006.
[15] P. Alvarado, A. Berner, and S. Akyol, "Combination of High-Level Cues in Unsupervised Single Image Segmentation Using Bayesian Belief Networks," *Proc. Int'l Conf. Imaging Science, Systems, and Technology,* vol. 2, pp. 675-681, 2002.
[16] S. Zheng, Z. Tu, and A. Yuille, "Detecting Object Boundaries Using Low-, Mid-, and High-Level Information," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2007.
[17] B. Leibe, A. Leonardis, and B. Schiele, "Combined Object Categorization and Segmentation with an Implicit Shape Model," *Proc. European Conf. Computer Vision Workshop Statistical Learning in Computer Vision,* pp. 17-32, 2004.
[18] J. Borges, J. Bioucas Dias, and A. Marcal, "Bayesian Hyperspectral Image Segmentation with Discriminative Class Learning," *Proc. Third Iberian Conf. Pattern Recognition and Image Analysis,* pp. I: 22-29, 2007.
[19] T. Hosaka, T. Kobayashi, and N. Otsu, "Image Segmentation Using MAP-MRF Estimation and Support Vector Machine," *Interdisciplinary Information Sciences,* vol. 13, no. 1, pp. 33-42, 2007.
[20] J. Winn and N. Jojic, "Locus: Learning Object Classes with Unsupervised Segmentation," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 756-763, 2005.
[21] S. Khan and M. Shah, "Object Based Segmentation of Video Using Color, Motion and Spatial Information," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 746-751, 2001.
[22] Y.-W. Tai, J. Jia, and C.-K. Tang, "Soft Color Segmentation and Its Applications," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 9, pp. 1520-1537, Sept. 2007.
[23] E. Borenstein and J. Malik, "Shape Guided Object Segmentation," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 969-976, 2006.
[24] Z. Tu, X. Chen, A.L. Yuille, and S.-C. Zhu, "Image Parsing: Unifying Segmentation, Detection, and Recognition," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 18-25, 2003.
[25] Z. Tu, C. Narr, P. Dollar, I. Dinov, P. Thompson, and A. Toga, "Brain Anatomical Structure Segmentation by Hybrid Discriminative/Generative Models," *IEEE Trans. Medical Imaging,* vol. 27, no. 4, pp. 495-508, Apr. 2008.
[26] S.Z. Li, *Markov Random Field Modeling in Image Analysis.* Springer, 2001.
[27] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," *Proc. Int'l Conf. Machine Learning,* pp. 282-289, 2001.
[28] S.L. Lauritzen, *Graphical Models.* Oxford Univ. Press, 1996.
[29] F.V. Jensen, *Bayesian Networks and Decision Graphs.* Springer-Verlag, 2001.
[30] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan-Kaufmann Publishers, 1988.
[31] C.M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2006.
[32] E. Borenstein, E. Sharon, and S. Ullman, "Combining Top-Down and Bottom-Up Segmentation," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition Workshop Perceptual Organization in Computer Vision,* p. 46, 2004.
[33] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation," *Proc. European Conf. Computer Vision,* pp. 1-15, 2006.
[34] M. Everingham, "The VOC2006 Database," Univ. of Oxford, http://www.pascal-network.org/challenges/VOC/databases.html#VOC2006, 2010.
[35] C. Bouman and M. Shapiro, "A Multiscale Random Field Model for Bayesian Image Segmentation," *IEEE Trans. Image Processing,* vol. 3, no. 2, pp. 162-177, Mar. 1994.
[36] H. Cheng and C. Bouman, "Multiscale Bayesian Segmentation Using a Trainable Context Model," *IEEE Trans. Image Processing,* vol. 10, no. 4, pp. 511-525, Apr. 2001.
[37] R. Wilson and C. Li, "A Class of Discrete Multiresolution Random Fields and Its Application to Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 1, pp. 42-56, Jan. 2003.
[38] W. Irving, P. Fieguth, and A. Willsky, "An Overlapping Tree Approach to Multiscale Stochastic Modeling and Estimation," *IEEE Trans. Image Processing,* vol. 6, no. 11, pp. 1517-1529, Nov. 1997.
[39] S. Kumar and M. Hebert, "Discriminative Random Fields," *Int'l J. Computer Vision,* vol. 68, no. 2, pp. 179-201, 2006.
[40] S. Kumar and M. Hebert, "A Hierarchical Field Framework for Unified Context-Based Classification," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 1284-1291, 2005.
[41] A. Kapoor and J. Winn, "Located Hidden Random Fields: Learning Discriminative Parts for Object Detection," *Proc. European Conf. Computer Vision,* pp. 302-315, 2006.
[42] J. Winn and J. Shotton, "The Layout Consistent Random Field for Recognizing and Segmenting Partially Occluded Objects," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 37-44, 2006.
[43] D. Hoiem, C. Rother, and J. Winn, "3D LayoutCRF for Multi-View Object Class Recognition and Segmentation," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2007.
[44] S. Sarkar and K.L. Boyer, "Integration, Inference, and Management of Spatial Information Using Bayesian Networks: Perceptual Organization," *IEEE Trans. Pattern Analysis and Machine Intelligence,* special section on probabilistic reasoning, vol. 15, no. 3, pp. 256-274, Mar. 1993.
[45] S.C. Zhu and A. Yuille, "Region Competition: Unifying Snake/Balloon, Region Growing and Bayes/MDL/Energy for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 9, pp. 884-900, Sept. 1996.

[46] S. Todorovic and M. Nechyba, "Dynamic Trees for Unsupervised Segmentation and Matching of Image Regions," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 11, pp. 1762-1777, Nov. 2005.

[47] R. Huang, V. Pavlovic, and D. Metaxas, "A Graphical Model Framework for Coupling MRFs and Deformable Models," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 739-746, 2004.

[48] C.H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, "Segmenting Brain Tumors with Conditional Random Fields and Support Vector Machines," *Proc. Computer Vision for Biomedical Image Applications: Current Techniques and Future Trends (within Int'l Conf. Computer Vision),* pp. 469-478, 2005.

[49] F. Liu, D. Xu, C. Yuan, and W. Kerwin, "Image Segmentation Based on Bayesian Network-Markov Random Field Model and its Application on in vivo Plaque Composition," *Proc. IEEE Int'l Symp. Biomedical Imaging,* pp. 141-144, 2006.

[50] V. Murino, C.S. Regazzoni, and G. Vernazza, "Distributed Propagation of A-Priori Constraints in a Bayesian Network of Markov Random Fields," *IEE Proc. I Comm., Speech and Vision,* vol. 140, no. 1, pp. 46-55, Feb. 1993.

[51] M.P. Kumar, P.H.S. Torr, and A. Zisserman, "OBJ CUT," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 18-25, 2005.

[52] G.E. Hinton, S. Osindero, and K. Bao, "Learning Causally Linked Markov Random Fields," *Proc. 10th Int'l Workshop Artificial Intelligence and Statistics,* 2005.

[53] B. Sumengen and B.S. Manjunath, "Edgeflow-Driven Variational Image Segmentation: Theory and Performance Evaluation," technical report, Univ. of California, Santa Barbara, http://barissumengen.com/seg/, 2005.

[54] S. Kumar and M. Hebert, "Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 1150-1157, 2003.

[55] S.V.N. Vishwanathan, N.N. Schraudolph, M.W. Schmidt, and K.P. Murphy, "Accelerated Training of Conditional Random Fields with Stochastic Gradient Methods," *Proc. Int'l Conf. Machine Learning,* vol. 148, pp. 969-976, 2006.

[56] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic Solution of Ill-Posed Problems in Computational Vision," *J. Am. Statistical Assoc.,* vol. 82, pp. 76-89, 1987.

[57] J. Yedidia, W. Freeman, and Y. Weiss, "Understanding Belief Propagation and Its Generalizations," *Proc. Int'l Joint Conf. Artificial Intelligence,* 2001.

[58] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Fourth Alvey Vision Conf.,* pp. 147-152, 1988.

[59] M. Pradhan, M. Henrion, G.M. Provan, B.D. Favero, and K. Huang, "The Sensitivity of Belief Networks to Imprecise Probabilities: An Experimental Investigation," *Artificial Intelligence,* vol. 85, nos. 1/2, pp. 363-397, 1996.

[60] F. Kschischang, B.J. Frey, and H.-A. Loeliger, "Factor Graphs and the Sum-Product Algorithm," *IEEE Trans. Information Theory,* vol. 47, no. 2, pp. 498-519, Feb. 2001.

[61] B.J. Frey, "Extending Factor Graphs so as to Unify Directed and Undirected Graphical Models," *Proc. 19th Conf. Uncertainty in Artificial Intelligence,* pp. 257-264, 2003.

[62] J. Park, "Using Weighted Max-Sat Engines to Solve MPE," *Proc. 18th Nat'l Conf. Artificial Intelligence,* pp. 682-687, 2002.

[63] F. Hutter, H.H. Hoos, and T. Stutzle, "Efficient Stochastic Local Search for MPE Solving," *Proc. Int'l Joint Conf. Artificial Intelligence,* pp. 169-174, 2005.

[64] T. Cour and J. Shi, "Recognizing Objects by Piecing Together the Segmentation Puzzle," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2007.

[65] A. Levin and Y. Weiss, "Learning to Combine Bottom-Up and Top-Down Segmentation," *Proc. European Conf. Computer Vision,* pp. 581-594, 2006.

[66] L. Zhu, C. Lin, H. Huang, Y. Chen, and A.L. Yuille, "Unsupervised Structure Learning: Hierarchical Recursive Composition, Suspicious Coincidence and Competitive Exclusion," *Proc. European Conf. Computer Vision,* pp. 759-773, 2008.

[67] E. Borenstein and S. Ullman, "Learning to Segment," *Proc. European Conf. Computer Vision,* pp. 1-8, 2004.

[68] L. Yang, P. Meer, and D.J. Foran, "Multiple Class Segmentation Using a Unified Framework over Mean-Shift Patches," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2007.

[69] M. Varma and A. Zisserman, "A Statistical Approach to Texture Classification from Single Images," *Int'l J. Computer Vision,* vol. 62, nos. 1/2, pp. 61-81, 2005.

[70] Z. Tu, "Auto-Context and Its Application to High-Level Vision Tasks," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2008.

Lei Zhang received the BS and MS degrees in electrical engineering from Tsinghua University, China, in 1999 and 2002, respectively. He recently received the PhD degree in electrical engineering from Rensselaer Polytechnic Institute in May 2009. His research focuses on machine learning and its application to computer vision problems. His current research is primarily about developing probabilistic graphical models such as Bayesian Network, dynamic Bayesian Network, condition random field, factor graph, chain graph, etc. He has applied different graphical models to several computer vision problems, including image segmentation, upper body tracking, facial expression recognition, etc. He is a member of the IEEE and a full member of Sigma Xi, the Scientific Research Society.

Qiang Ji received the PhD degree in electrical engineering from the University of Washington. He is currently a professor with the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute (RPI). He is also a program director at the US National Science Foundation (NSF), managing NSF's computer vision and machine learning programs. He has also held teaching and research positions with the Beckman Institute at the University of Illinois at Urbana-Champaign, the Robotics Institute at Carnegie Mellon University, the Department of Computer Science at the University of Nevada at Reno, and the US Air Force Research Laboratory. He currently serves as the director of the Intelligent Systems Laboratory (ISL) at RPI. His research interests are in computer vision, pattern recognition, and probabilistic machine learning and their applications in various fields. He has published more than 150 papers in peer-reviewed journals and conferences. His research has been supported by major governmental agencies including the NSF, US National Institutes of Health, US Defense Advanced Research Projects Agency (DARPA), US Office of Naval Research, US Army Research Office, and US Air Force Office of Scientific Research, as well as by major companies including Honda and Boeing. He is an editor on several computer vision and pattern recognition related journals and he has served as a program chair, technical area chair, and program commitee member for numerous international conferences/workshops. He is a senior member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.