Image Modeling using Tree Structured Conditional Random Fields

Pranjal Awasthi
IBM India Research Lab*
New Delhi
prawasth@in.ibm.com

Aakanksha Gagrani
Dept. of CSE
IIT Madras
aksgag@cse.iitm.ernet.in

Balaraman Ravindran
Dept. of CSE
IIT Madras
ravi@cs.iitm.ernet.in

Abstract

In this paper we present a discriminative framework based on conditional random fields for stochastic modeling of images in a hierarchical fashion. The main advantage of the proposed framework is its ability to incorporate a rich set of interactions among the image sites. We achieve this by inducing a hierarchy of hidden variables over the given label field. The proposed tree like structure of our model eliminates the need for a huge parameter space and at the same time permits the use of exact and efficient inference procedures based on belief propagation. We demonstrate the generality of our approach by applying it to two important computer vision tasks, namely image labeling and object detection. The model parameters are trained using the contrastive divergence algorithm. We report the performance on real world images and compare it with the existing approaches.

1 Introduction

Probabilistic modeling techniques have been used effectively in several computer vision related tasks such as image segmentation, classification, object detection etc. These tasks typically require processing information available at different levels of granularity. In this work we propose a probabilistic framework that can incorporate image information at various scales in a meaningful and tractable model. Our framework, like many of the existing image modeling approaches, is built on Random field theory. Random field theory models the uncertain information by encoding it in terms of a set $\mathcal{S}$ of random variables and an underlying graph structure $\mathcal{G}$. The graph structure represents the dependence relationship among these variables. We are interested in the problem of inferencing which can be stated as: Given the values taken by some subset $\mathbf{X} \in \mathcal{S}$, estimate the most probable values for another subset $\mathbf{Y} \in \mathcal{S}$. Conventionally, we define the variables in the set X as observations and the variables in the set Y as labels.

The information relevant for constructing an accurate model of the image can be broadly classified into two categories widely known as *bottom-up cues* and *top-down cues*.

Bottom-up cues refer to the statistical properties of individual pixels or a group of pixels derived from the outputs of various filters. *Top-down cues* refer to the contextual features present in the image. This information represents the interactions that exist among the labels in the set $\mathbf{Y}$. These *top-down cues* need to be captured at different levels of scale for different tasks as described in [Kumar and Hebert, 2005]. Bottom-up cues such as color and texture features of the image are insufficient and ambiguous for discriminating the pixels with similar spectral signatures. With reference to Figure 1 taken from the Corel image database, it can be observed that {rhino, land} and {sea, sky} pixels are closely related in terms of their spectral properties. However, an expanded context window can resolve the ambiguity. It also obviates the improbable configurations like the occurrence of a boundary between snow and water pixels. Further enlarging the context window captures expected global relationships such as 1) polar bear surrounded by snow and 2) sky at the top of the image while water at the bottom. Any system that fails to capture them is likely to make a large number of errors on difficult test images. In this work we try to model these relationships by proposing a discriminative hierarchical framework.

Figure 1: Contextual features at multiple scales play a vital role in discrimination when the use of local feature fails to give a conclusive answer.

1.1 Related Work

In the past, [Geman and Geman, 1984] have proposed Markov Random Field (MRF) based models for joint mod-

---

*Part of this work was done while the author was at IIT Madras.

eling of the variables **X** and **Y**. The generative nature of MRFs restricts the underlying graph structure to a simple two dimensional lattice which can capture only local relationships among neighboring pixels, e.g. within a 4×4 or a 8×8 window. The computational complexity of the training and inference procedures for MRFs rules out the possibility of using a larger window size. Multiscale Random Fields (MSRFs) [Bouman and Shapiro, 1994] try to remove this limitation by capturing the interactions among labels in a hierarchical framework. Similar approach has been proposed by [Feng *et al.*, 2002] using Tree Structured Bayesian Networks (TSBNs). These approaches have the ability to capture long range label relationships. In-spite of the advantage offered by the hierarchical nature, these methods suffer from the same limitations of the generative models, namely,

- Strong independence assumptions among the observations **X**.

- Introduction of a large number of parameters by modeling the joint probability $P(\mathbf{Y}, \mathbf{X})$.

- Need for large training data.

Recently, with the introduction of Conditional Random Fields (CRFs), the use of discriminative models for classification tasks has become popular [Lafferty *et al.*, 2001]. CRFs offer a lot of advantages over the generative approaches by directly modeling the conditional probability $P(\mathbf{Y} \,|\, \mathbf{X})$. Since no assumption is made about the underlying structure of the observations **X**, the model is able to incorporate a rich set of non-independent overlapping features of the observations. In the context of images, various authors have successfully applied CRFs for classification tasks and have reported significant improvement over the MRF based generative models [He *et al.*, 2004; Kumar and Hebert, 2003].

Discriminative Random Fields (DRFs) which closely resemble CRFs have been applied for the task of detecting manmade structures in real world images [Kumar and Hebert, 2003]. Though the model outperforms traditional MRFs, it is not strong enough to capture long range correlations among the labels due to the rigid lattice based structure which allows for only pairwise interactions. [He *et al.*, 2004] propose Multiscale Conditional Random Fields (mCRFs) which are capable of modeling interactions among the labels over multiple scales. Their work consists of a combination of three different classifiers operating at local, regional and global contexts respectively. However, it suffers from two main drawbacks

- Including additional classifiers operating at different scales into the mCRF framework introduces a large number of model parameters.

- The model assumes conditional independence of hidden variables given the label field.

Though a number of generative hierarchical models have been proposed, similar work involving discriminative frameworks is limited [Kumar and Hebert, 2005]. In this paper we propose a Tree Structured Conditional Random Field (TCRF) for modeling images. Our results demonstrate that TCRF is able to capture long range label relationships and is robust enough to be applied to different classification tasks. The primary contribution of our work is a method which combines the advantages offered by hierarchical models and discriminative approaches in a single framework. The rest of the paper is organized as follows: We formally present our framework in Section 2. In the section we also discuss the graph structure and the form of the probability distribution. Training and inference for a TCRF are discussed in sections 3 and 4 respectively. We present the results of our model on two different tasks in section 5. A few possible extensions to our framework are discussed in section 6. We finally conclude in section 7.

2 Tree Structured Conditional Random Field

Let **X** be the observations and **Y** the corresponding labels. Before presenting our framework we first state the definition of Conditional Random Fields as given by [Lafferty *et al.*, 2001].

2.1 CRF Definition

Let $G = (V, E)$ be a graph such that **Y** is indexed by the vertices of G. Then $(\mathbf{X}, \mathbf{Y})$ is a conditional random field if, when conditioned on **X**, the random variables $y_i \in$ **Y** obey the Markov property with respect to the graph: $p(y_i \,|\, \mathbf{X}, y_j, i \neq j) = p(y_i \,|\, \mathbf{X}, y_j, i \sim j)$. Here $i \sim j$ implies that y_i and y_j are neighbors in G.

By applying the Hammersley-Clifford theorem [Li, 2001], the conditional probability $p(\mathbf{Y} \,|\, \mathbf{X})$ factors into a product of potential functions. These real valued functions are defined over the cliques of the graph G.

$$p(\mathbf{Y} \,|\, \mathbf{X}) = \frac{1}{Z} \prod_{c \in C} \psi_c(\mathbf{y}_c \,, \mathbf{X}) \tag{1}$$

In the above equation C is the set of cliques of the graph G, $\mathbf{y}_c$ is the set of variables in **Y** which belong to the clique c and Z is a normalization factor.

2.2 TCRF Graph Structure

Consider a 2×2 neighborhood of pixels as shown in Figure 2. One way to model the association between the labels of these pixels is to introduce a weight vector for every edge (u, v) which represents the compatibility between the labels of the nodes u and v (Figure 2a). An alternative is to introduce a hidden variable H which is connected to all the 4 nodes. For every value which variable H takes, it induces a probability distribution over the labels of the nodes connected to it (Figure 2b).
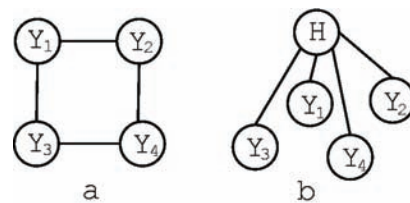

Figure 2: Two different ways of modeling interactions among neighboring pixels.

Dividing the whole image into regions of size $m \times m$ and introducing a hidden variable for each of them gives a layer of hidden variables over the given label field. In such a configuration each label node is associated with a hidden variable in the layer above. The main advantage of following such an approach is that long range correlations among non-neighboring pixels can be easily modeled as associations among the hidden variables in the layer above. Another attractive feature is that the resulting graph structure induced is a tree which allows inference to be carried out in a time linear in the number of pixels [Felzenszwalb and Huttenlocher, 2004]. Motivated by this, we introduce multiple layers of hidden variables over the given label field. Each layer of hidden variables tries to capture label relationships at a different level of scale by affecting the values of the hidden variables in the layer below. In our work we take the value of m to be equal to 2. The quadtree structure of our model is shown in Figure 3.
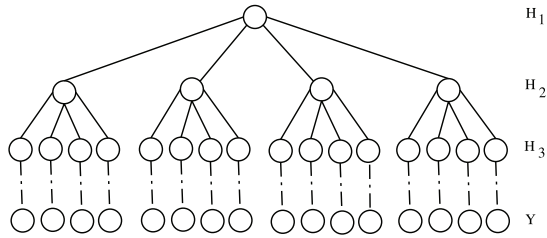

Figure 3: A tree structured conditional random field.

Formally, let $\mathbf{Y}$ be the set of label nodes, $\mathbf{X}$ the set of observations and $\mathbf{H}$ be the set of hidden variables. We call $(\{\mathbf{Y}, \mathbf{H}\}, \mathbf{X})$ a TCRF if the following conditions hold:

1. There exists a connected acyclic graph $G = (V, E)$ whose vertices are in one-to-one correspondence with the variables in $\mathbf{Y} \cup \mathbf{H}$, and whose number of edges $|E| = |V| - 1$.

2. The node set V can be partitioned into subsets $(V_1, V_2, \ldots, V_L)$ such that $\bigcup_i V_i = V$, $V_i \cap V_j = \emptyset$ and the vertices in V_L correspond to the variables in $\mathbf{Y}$.

3. $deg(v) = 1, \quad \forall v \in V_L$.

Similar to CRFs, the conditional probability $P(\mathbf{Y}, \mathbf{H} \mid \mathbf{X})$ of a TCRF factors into a product of potential functions. We next describe the form of these functions.

2.3 Potential Functions

Since the underlying graph for a TCRF is a tree, the set of cliques C corresponds to nodes and edges of this graph. We define two kinds of potentials over these cliques:

Local Potential
This is intended to represent the influence of the observations $\mathbf{X}$ on the label nodes $\mathbf{Y}$. For every $\mathbf{y_i} \in \mathbf{Y}$ this function takes the form $\exp(\Gamma_i(\mathbf{y_i}, \mathbf{X}))$. $\Gamma_i(\mathbf{y_i}, \mathbf{X})$ can be defined as

$$\Gamma_i(\mathbf{y_i}, \mathbf{X}) = \mathbf{y_i}^T \mathbf{W} \mathbf{f}_i(\mathbf{X}) \tag{2}$$

Let $\mathcal{L}$ be the number of classes and $\mathcal{F}$ be the number of image statistics for each block. Then $\mathbf{y}_i$ is a vector of length

$\mathcal{L}$, $\mathbf{W}$ is a matrix of weights of size $\mathcal{L} \times \mathcal{F}$ estimated from the training data and $\mathbf{f}(.)$ is a transformation (possibly non-linear) applied to the observations. Another approach is to take $\Gamma_i(\mathbf{y_i}, \mathbf{X}) = \log P(\mathbf{y}_i \mid \mathbf{X})$, where $P(\mathbf{y}_i \mid \mathbf{X})$ is the probability estimate output by a separately trained local classifier [He *et al.*, 2006].

Edge Potential
This is defined over the edges of the tree. Let $\phi^{t,b}$ be a matrix of weights of size $\mathcal{L} \times \mathcal{L}$ which represents the compatibility between the values of a hidden variable at level t and its bth neighbor at level $t+1$. For our quad-tree structure $b = 1, 2, 3$ or 4. Let L denote the number of levels in the tree, with the root node at level 1 and the image sites at level L and $\mathbf{H}_t$ denote the set of hidden variables present at level t. Then for every edge $(\mathbf{h}_i, \mathbf{y}_j)$ such that $\mathbf{h}_i \in \mathbf{H}_{L-1}$ and $\mathbf{y}_j \in \mathbf{Y}$ the edge potential can be defined as $\exp(\mathbf{h}_i^T \phi^{L-1,b} \mathbf{y}_j)$. In a similar manner the edge potential between the node $\mathbf{h}_i$ at level t and its bth neighbor $\mathbf{h}_j$ at level $t + 1$ can be represented as $\exp(\mathbf{h}_i^T \phi^{t,b} \mathbf{h}_j)$.

The overall joint probability distribution which is a product of the potentials defined above can be written as

$$P(\mathbf{Y}, \mathbf{H} \mid \mathbf{X}) \quad \propto \quad \exp(\sum_{\mathbf{y}_i \in \mathbf{Y}} \mathbf{y}_i^T \mathbf{W} \mathbf{f}_i(\mathbf{X}) \tag{3}$$
$$+ \sum_{\substack{\mathbf{y}_j \in \mathbf{Y} \\ (\mathbf{y}_j, \mathbf{h}_i) \in E}} \sum_{\mathbf{h}_i \in \mathbf{H}_{L-1}} \mathbf{h}_i^T \phi^{L-1,b} \mathbf{y}_j$$
$$+ \sum_{t=1}^{L-2} \sum_{\mathbf{h}_i \in \mathbf{H}_t} \sum_{\substack{\mathbf{h}_j \in \mathbf{H}_{t+1} \\ (\mathbf{h}_i, \mathbf{h}_j) \in E}} \mathbf{h}_i^T \phi^{t,b} \mathbf{h}_j)$$

3 Parameter Estimation

Given a set of training images $\mathcal{T} = \{(\mathbf{Y}_1, \mathbf{X}_1) \ldots (\mathbf{Y}_N, \mathbf{X}_N)\}$ we want to estimate the parameters $\mathbf{\Theta} = \{\mathbf{W}, \mathbf{\Phi}\}$. A standard way to do this is to maximize the conditional log-likelihood of the training data. This is equivalent to minimizing the Kullback-Leibler (KL) divergence between the model distribution $P(\mathbf{Y}|\mathbf{X}, \mathbf{\Theta})$ and the empirical distribution $Q(\mathbf{Y}|\mathbf{X})$ defined by the training set. The above approach, requires calculating expectations under the model distribution. These expectations are approximated using Markov Chain Monte Carlo (MCMC) methods. As pointed by [Hinton, 2002], these methods suffer from two main drawbacks

1. The Markov chain takes a long time to reach the equilibrium distribution.

2. The estimated gradients tend to be noisy, i.e. they have a high variance.

Contrastive Divergence (CD) is an approximate learning method which overcomes this problem by minimizing a different objective function [Hinton, 2002]. In this work we use CD learning for estimating the parameters.

3.1 Contrastive Divergence Learning

Let P be the model distribution and Q^0 be the empirical distribution of label variables. CD learning tries to minimize the

contrastive divergence which is defined as

$$\mathcal{D} = Q^0 \parallel P - Q^1 \parallel P \qquad (4)$$

where $Q \parallel P = \sum_{\mathbf{y} \in \mathbf{Y}} Q(\mathbf{y}) \log \frac{Q(\mathbf{y})}{P(\mathbf{y})}$ is the Kullback-Leibler divergence between Q & P. Q^1 is the distribution defined by the one step reconstruction of the training data vectors [Hinton, 2002]. For any weight w_{ij},

$$\Delta w_{ij} = -\eta \frac{\partial \mathcal{D}}{\partial w_{ij}} \qquad (5)$$

where η is the learning rate. Let $\mathbf{Y}^1$ be the one step reconstruction of training set $\mathbf{Y}^0$. The weight update equations specific to our model are

$$\Delta w_{uv} = \eta [\sum_{\mathbf{y}_i \in \mathbf{Y}^0} y_i^u f_i^v(\mathbf{X}) - \sum_{\mathbf{y}_i \in \mathbf{Y}^1} y_i^u f_i^v(\mathbf{X})] \qquad (6)$$

$$\Delta \phi_{uv}^{L-1,b} = \eta \Big[\sum_{\substack{\mathbf{y}_j \in \mathbf{Y}^0 \\ (\mathbf{y}_j, \mathbf{h}_i) \in \mathbf{E}}} \sum_{\mathbf{h}_i \in \mathbf{H}_{L-1}} \mathcal{E}_{P(\mathbf{h}_i \mid \mathbf{Y}^0)}[h_i^u y_j^v] \qquad (7)$$
$$- \sum_{\substack{\mathbf{y}_j \in \mathbf{Y}^1 \\ (\mathbf{y}_j, \mathbf{h}_i) \in \mathbf{E}}} \sum_{\mathbf{h}_i \in \mathbf{H}_{L-1}} \mathcal{E}_{P(\mathbf{h}_i \mid \mathbf{Y}^1)}[h_i^u y_j^v] \Big]$$

$$\Delta \phi_{uv}^{t,b} = \eta \Big[\sum_{\substack{\mathbf{h}_i \in \mathbf{H}_t \\ (\mathbf{h}_i, \mathbf{h}_j) \in \mathbf{E}}} \sum_{\mathbf{h}_j \in \mathbf{H}_{t+1}} \mathcal{E}_{P(\mathbf{h}_i, \mathbf{h}_j \mid \mathbf{Y}^0)}[h_i^u h_j^v] \qquad (8)$$
$$- \sum_{\substack{\mathbf{h}_i \in \mathbf{H}_t \\ (\mathbf{h}_i, \mathbf{h}_j) \in \mathbf{E}}} \sum_{\mathbf{h}_j \in \mathbf{H}_{t+1}} \mathcal{E}_{P(\mathbf{h}_i, \mathbf{h}_j \mid \mathbf{Y}^1)}[h_i^u h_j^v] \Big]$$

4 Inference

The problem of inference is to find the optimal label field $\mathbf{Y}^* = argmax_{\mathbf{Y}} P(\mathbf{Y}|\mathbf{X})$. This can be done using the maximum posterior marginals (MPM) criterion which minimizes the expected number of incorrectly labeled sites. According to the MPM criterion, the optimal label for pixel i, $\mathbf{y}_i^*$ is given as

$$\mathbf{y}_i^* = argmax_{\mathbf{y}_i} P(\mathbf{y}_i \mid \mathbf{X}), \quad \forall \mathbf{y}_i \in \mathbf{y} \qquad (9)$$

Due to the tree structure of our model the probability value $P(\mathbf{y}_i|\mathbf{X})$ can be exactly calculated using Belief Propagation [Pearl, 1988].

5 Experiments and Results

In order to evaluate the performance of our model, we applied it to two different classification tasks, namely, object detection and image labeling. Our main aim was to demonstrate the ability of TCRF in capturing the label relationships. We have also compared our approach with the existing techniques both qualitatively and quantitatively. The predicted label for each block is assigned to all the pixels contained within it. This is done in order to compare the performance against models like mCRF which operate directly at the pixel level. We use a common set of local features as described next.

5.1 Feature Extraction

Visual contents of an image such as color, texture, and spatial layout are vital discriminatory features when considering real world images. For extracting the color information we calculate the first three color moments of an image patch in the $CIEL^*a^*b^*$ color space.

We incorporate the entropy of a region to account for its randomness. A uniform image patch has a significantly lower entropy value as compared to a non-uniform patch. Entropy is calculated as shown in Equation 10

$$\mathcal{E} = - \sum_p (p * log(p)) \qquad (10)$$

where, p contains the histogram counts. Texture is extracted using the discrete wavelet transform [Mallat, 1996]. Spatial layout is accounted for by including the coordinates of every image site. This in all accounts for a total of 20 statistics corresponding to one image region.

5.2 Object Detection

We apply our model to the task of detecting man-made structures in a set of natural images from the Corel image database. Each image is 256×384 pixels. We divide each image into blocks of size 8×12. Each block is labeled as either *structured* or *unstructured*. These blocks correspond to the leaf nodes of the TCRF. In order to compare our performance with DRFs, we use the same set of 108 training and 129 testing images as used by [Kumar and Hebert, 2003]. We also implement a simple Logistic Regression (LR) based classifier in order to demonstrate the performance improvement obtained by capturing label relationships. As shown in Table 1, TCRF achieves significant improvement over the LR based classifier which considers only local image statistics into account. Figure 4 shows some the results obtained by our model.

Table 1: Classification Accuracy for detecting man-made structures calculated for the test set containing 129 images.

Model	Classification Accuracy (%)
LR	69.82
DRF	72.54[1]
TCRF	85.83

5.3 Image Labeling

We now demonstrate the performance of our model on labeling real world images. We took a set of 100 images consisting of wildlife scenes from the Corel database. The dataset contains a total of 7 class labels: *rhino/hippo, polar bear, vegetation, sky, water, snow and ground*. This is a data set with high variability among images. Each image is 128×192 pixels. We divide each image into blocks of size 4×6. A total of 60 images were chosen at random for training. The remaining

---

[1]This score as reported by [Kumar and Hebert, 2003] was obtained by using a different region size.

40 images were used for testing. We compare our results with that of a logistic classifier and our own implementation of the mCRF [He *et al.*, 2004]. It can be observed from Table 2 that TCRF significantly improves over the logistic classifier by modeling the label interactions. Our model also outperforms the mCRF on the given test set. Qualitative results on some of the test images are shown in Figure 5.

Table 2: Classification accuracy for the task of image labeling on the Corel data set. A total of 40 testing images were present each of size 128×192.

Model	Classification Accuracy (%)
LR	65.03
mCRF	74.3
TCRF	77.76

6 Discussion

The framework proposed in this paper falls into the category of multiscale models which have been successfully applied by various authors [Bouman and Shapiro, 1994; Feng *et al.*, 2002]. The novelty of our work is that we extend earlier hierarchical approaches to incorporate them into a discriminative framework based on conditional random fields. There are several immediate extensions possible to the basic model presented in this work. One of them concerns with the "blocky" nature of the labelings obtained. The reason for this is the fixed region size which is used in our experiments. This behavior is typical of tree based models like TSBNs. One solution to this problem is to directly model image pixels as the leaves of the TCRF. This would require increasing the number of levels of the tree. Although the number of model parameters increase only linearly with the number of levels, training and inference times can increase exponentially with the number of levels. Recently, He *et al.* have reported good results by applying their model over a super-pixelized representation of the image [He *et al.*, 2006]. This approach reduces the number of image sites and at the same time does not define rigid region sizes. We are currently exploring the possibility of incorporating super-pixelization into the TCRF framework.

7 Conclusions

We have presented TCRF, a novel hierarchical model based on CRFs for incorporating contextual features at several levels of granularity. The discriminative nature of CRFs combined with the tree like structure gives TCRFs several advantages over other proposed multiscale models. TCRFs yield a general framework which can be applied to a variety of image classification tasks. We have empirically demonstrated that a TCRF outperforms existing approaches on two such tasks. In future we wish to explore other training methods which can handle larger number of parameters. Further we need to study how our framework can be adapted for other image classification tasks.

References

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

[Felzenszwalb and Huttenlocher, 2004] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient belief propagation for early vision. In *CVPR (1)*, pages 261–268, 2004.

[Feng *et al.*, 2002] Xiaojuan Feng, Christopher K. I. Williams, and Stephen N. Felderhof. Combining belief networks and neural networks for scene segmentation. *IEEE Trans. Pattern Anal. Mach. Intell*, 24(4):467–483, 2002.

[Geman and Geman, 1984] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, 1984.

[He *et al.*, 2004] Xuming He, Richard S. Zemel, and Miguel Á. Carreira-Perpiñán. Multiscale conditional random fields for image labeling. In *CVPR (2)*, pages 695–702, 2004.

[He *et al.*, 2006] Xuming He, Richard S. Zemel, and Debajyoti Ray. Learning and incorporating top-down cues in image segmentation. In *ICCV*, 2006.

[Hinton, 2002] Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002.

[Kumar and Hebert, 2003] Sanjiv Kumar and Martial Hebert. Discriminative fields for modeling spatial dependencies in natural images. In *NIPS*. MIT Press, 2003.

[Kumar and Hebert, 2005] Sanjiv Kumar and Martial Hebert. A hierarchical field framework for unified context-based classification. In *ICCV*, pages 1284–1291. IEEE Computer Society, 2005.

[Lafferty *et al.*, 2001] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, pages 282–289. Morgan Kaufmann, 2001.

[Li, 2001] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Comp. Sci. Workbench. Springer, 2001.

[Mallat, 1996] S. Mallat. Wavelets for a vision. *Proceedings of the IEEE*, 84(4):604–614, 1996.

[Pearl, 1988] J. Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann, 1988.
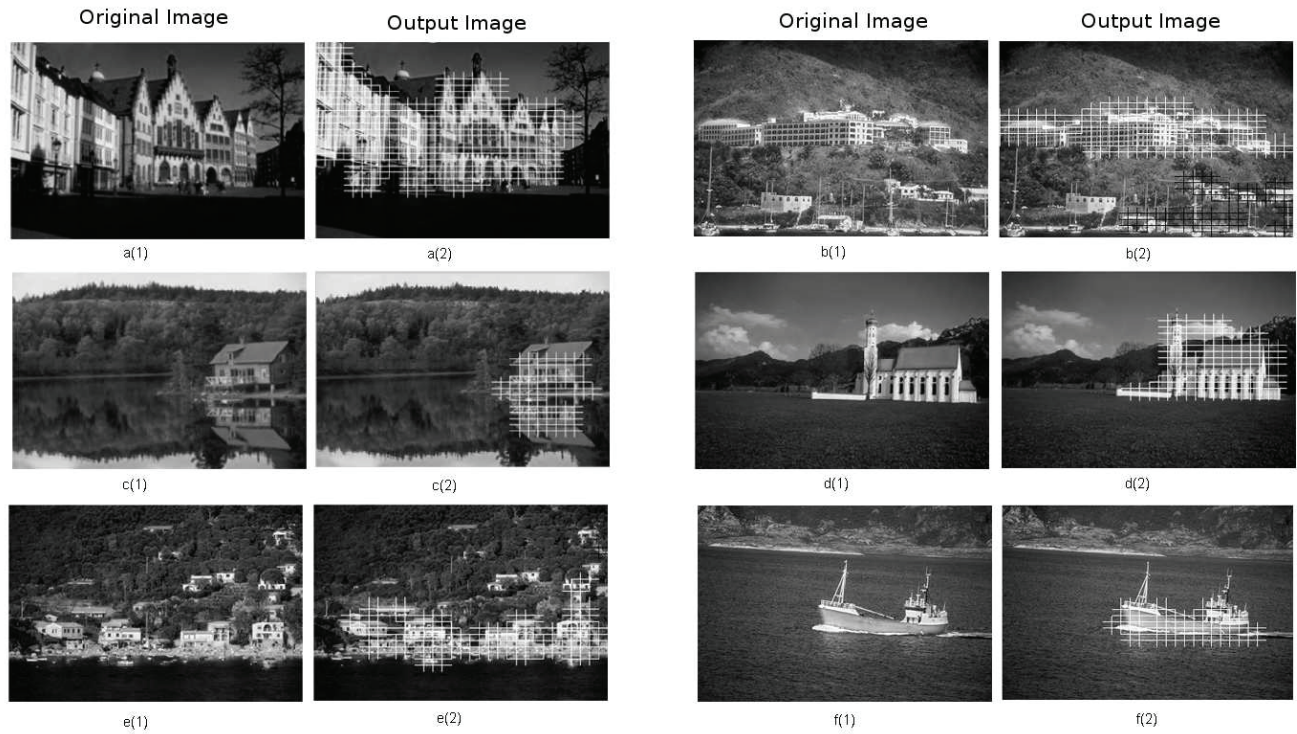
Figure 4: Man made structure detection results on the Corel image database. The grids in the output image represent the blocks classified as *structured*. Note that TCRF is able to give good results on difficult test images like b and e.
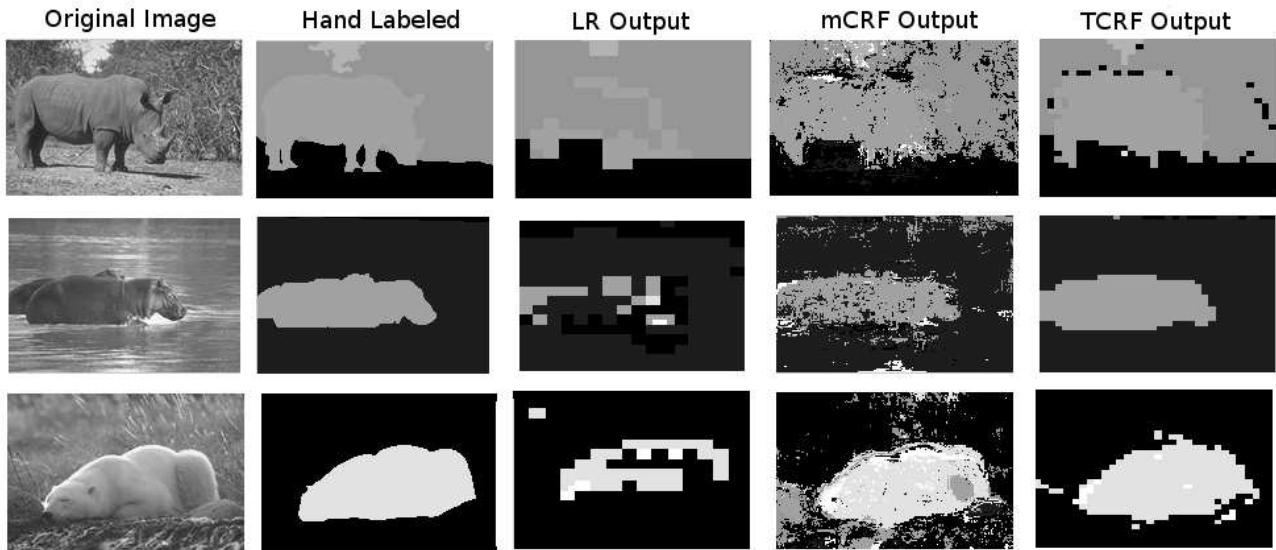

Figure 5: Image labeling results on the Corel test set. The TCRF achieves significant improvement over the logistic classifier by taking label relationships into account. It also gives good performance in cases where the mCRF performs badly as shown above.