Efficient Image Understanding Based on the Markov Random Field Model and Error Backpropagation Network

Il Y. Kim and Hyun S. Yang
Center for Artificial Intelligence Research
Dept. of Computer Science, KAIST
Taejon 305-701, KOREA

Abstract

Image labeling is a process of recognizing each segmented region properly exploiting the properties of the regions and the spatial relationships between regions. In some sense, image labeling is an optimization process of indexing regions using the constraints as to the scene knowledge. In this paper, we further investigate a method of efficiently labeling images using the Markov Random Field(MRF). MRF model is defined on the region adjacency graph and the labeling is then optimally determined using the simulated annealing. The MRF model parameters are automatically estimated using the error backpropagation network. We analyze the proposed method through experiments using the real natural scene images.

1 Introduction

Image understanding consists of image segmentation and image labeling in large. Image segmentation is a process of segmenting an image into a group of homogeneous regions whose characteristics such as graylevel, color, texture, range, etc. are similar while image labeling is a process of recognizing each segmented region exploiting the properties of the regions, spatial relatiohships between regions, knowledge as to the object models and the scene. In general, understanding images on 3-D scene is difficult because of the complexity and variability of the scene. Therefore, for reliable understanding of such an image, one must effectively exploit a set of knowledge as to the scene.

Among those various types of knowledge required for 3-D scene understanding, domain specific semantic knowledge is most frequently used for image labeling. Since semantic knowledge provides some constraints on labeling, the properties of the objects(or regions) are termed as unary constraints whereas the spatial relationships between objects(or regions) are termed as binary constraints. Thus, image labeling is a process of indexing images comprising many objects using the unary constraints on object properties and the binary constraints on the relationships between objects.

Markov Random Field(MRF) model provides a powerful mechanism for incorporating the spatial dependence of objects in the relative proximity of each other. And there is a very useful theorem that makes an explicit connection between the probabilities in the MRF formulation and the Gibbs distribution, which characterizes systems with energies and temparatures. So far, MRF model has been widely used to solve so called ill-posed problems in the computer vision areas such as image restoration [1], segmentation [2], [3], [4] and interpretation [5].

For the MRF-based image labeling, an image is first segmented into a set of disjoint regions and MRF is defined on the corresponding region adjacency graph. Figure 1 shows a synthetically generated segmented image and corresponding region adjacency graph.

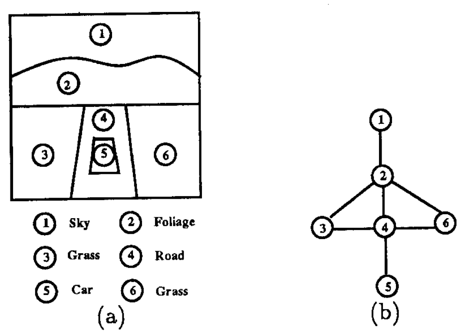

Figure 1: (a) Typical example of segmented image (b) corresponding region adjacency graph

Image labeling is then achieved through assigning object labels to the segmented regions, or nodes of the adjacency graph, using feature measurements and the domain knowledge such as unary and binary constraints. The labels are modeled as a MRF on the adjacency graph and image interpretation is then formulated as an optimization problem of maximizing the a posteriori probability of labeling given the feature measurements and the domain knowledge. Thus, by formulating image labeling problem as MRF model, the domain knowledge can be systematically represented in terms of the clique functions associated with cliques on the corresponding adjacency graph. Image labeling can then be achieved by a relatively efficient

0-8186-2910-X/92 $3.00 © 1992 IEEE

optimization algorithm, simulated annealing.

In general, parameters of MRF models are assumed to be known. But the values of these parameters determine both the distribution over the configuration sapce to which the system converges and the speed of convergence. Thus, estimating these parameters is very crucial in practice for the successful labeling. One way to estimate parameter values is using a set of examples in which data and desired solutions are given.

In this paper, we further investigate the MRF-based image labeling method. For estimating the parameters associated with the single-node clique functions, we implement the clique function as a neural network whose weights are learned from examples by error backpropagation method [6]. The best labeling that minimizes the MRF-based energy function is then obtained by simulated annealing algorithm [7].

2 MRF-based image labeling

2.1 Image labeling formulation

Now we briefly describe the underlying concepts associated with the MRF defined on the region adjacency graph and show how this can be applied to the image labeling problem. The MRF-based image labeling is previously formulated by Zhang, et.al.[5], and we will follow their formulation in basic scheme. But in designing clique functions we redefine a single-node clique function as a single parameter value, which is estimated by a multilayer neural network using error backpropagation algorithm.

Let $G=\{R,E\}$ be a graph, where

$$R = \{R_i : 1 \leq i \leq N\} \quad (1)$$

is a set of disjoint regions; and E is a set of edges connecting them. Suppose that there exist a neighborhood system on G, denoted by

$$N = \{n(R_i) : 1 \leq i \leq N\}, \quad (2)$$

where n(R_i) is a set of regions in R that are neighbors of R_i. Let $X = \{X_R,\ R \in R\}$ denote any family of random variables indexed by R, and Λ be a set of possible labels so that $X_R \in \Lambda$ for all R. Let Ω be the set of all possible *configurations:*

$$\Omega = \{\omega = (\lambda_{R_1}, ..., \lambda_{R_N}) : \lambda_{R_i} \in \Lambda, 1 \leq i \leq N\}. \quad (3)$$

As usual, the event $\{X_{R_1} = \lambda_{R_1}, ..., X_{R_N} = \lambda_{R_N}\}$ is abbreviated $\{X = \omega\}$. Then X is an MRF with respect to N if

$$\begin{gathered} P(X = \omega) > 0\ \forall \omega \in \Omega; \quad (4) \\ P(X_{R_i} = \lambda_{R_i} | X_{R_j} = \lambda_{R_j}, R_j \neq R_i) \\ = P(X_{R_i} = \lambda_{R_i} | X_{R_j} = \lambda_{R_j}, R_j \in n(R_i)) \end{gathered}$$

for every $R_i \in R$ and $(\lambda_{R_1}, ..., \lambda_{R_N}) \in \Omega$. And P[.] and P[. | .] are the joint and conditional pdf's, respectively. An important characteristics of the MRF model defined above is that its joint pdf has a general functional form, known as the Gibbs distribution, which is defined on the concept of cliques. So P($X = \omega$) can be written as

$$P(X = \omega) = Z^{-1} e^{-U(\omega)} \quad (5)$$

where U(ω) $= \sum_{c \in C} V_c(\omega)$. Here, U($\omega$) is called Gibb's energy function, $V_c(\omega)$ is called clique function defined on the corresponding clique c, and Z is the normalization constant.

Image interpretation problem is to find maximum a *posteriori* (MAP) estimate of X corresponding to the maximum of P($X = \omega$), and P($X = \omega$) can be maximized by minimizing Gibb's energy function, U(ω), which is the summation of $V_c(\omega)$'s. Therefore we should first define clique functions, $V_c(\omega)$ for the corresponding clique c.

2.2 Designing clique functions

A clique is a subset of R which contains either a single node or several nodes that are all neighbors of each other. In general, the optimal interpretation should be the one that is most consistent with the feature measurements and the domain knowledge. So a general principle for designing clique functions can be stated as follows; If the interpretation of the regions in a clique tends to be consistent with the feature measurements and the domain knowledge, the clique function decreases, resulting in a decrease in the energy function; otherwise, the clique function increase, resulting in an increase in the energy function.

We first consider the clique functions for single-node cliques. Let c be an arbitrary single-node clique with one region R. Let the corresponding clique function be defined by

$$V_c(\omega) = V_c(X_R = \lambda) = \alpha_\omega, \quad (6)$$

where λ is a label for R and α_ω is a parameter associated with ω. Usually, α_ω's are assumed known or estimated by some parametric functions. But this is not appropriate for the interpretation of the complex real natural scenes because there are so much variations on the input images and so it is difficult to pre-determine the parameter values or define estimation functions without invoking any other parameters. To overcome this, we implemented clique functions as multi-layer neural networks. The network produces the parameter values as its output and these parameter values are learned from examples by error backpropagation algorithm, detailed description of which will be found in the next section.

The clique functions for the two-node cliques can then be defined as follows. Let c be an arbitrary two-node clique with two adjacent regions, R_i and R_j in R. Then the clique function for the clique c can be defined as

$$V_c'(\omega) = V_c'(X_{R_i} = \lambda_i, X_{R_j} = \lambda_j) = \beta_\omega, \quad (7)$$

where λ_i and λ_j are labels for R_i and R_j, respectively. We set β_ω to 0 if λ_i and λ_j pair is a valid combination; otherwise, 1. Table 1 shows the example of the pairs of the labeled regions and their corresponding parameter values.

Table 1: Example of the pairs of two adjacent labeled regions and their corresponding parameter values

Pairs of the labeled regions	β_ω
sky, road	1
sky, centerline	1
sky, window	1
foliage, centerline	1
foliage, window	1
road, window	1
wall, centerline	1
centerline, window	1
the rest	0

3 Parameter estimation by error backpropagation

As discussed in the previous section, there are some parameters in the MRF model. These parameters are, in general, assumed to be known or determined heuristically. One way to estimate these parameter values is using a set of examples in which data and desired solutions are given. We exploit the error backpropagation network to estimate these parameters from examples.

Let W be a set of networks:

$$W = \{W_{\lambda_1}, ..., W_{\lambda_m} : \lambda_i \in \Lambda, 1 \leq i \leq m\}, \quad (8)$$

where m is the number of possible labels. Then we define again the single-node clique functions as

$$V_c(\omega) = V_c(X_R = \lambda) = O(W_\lambda; F(R)) \quad (9)$$

where W_λ is a network for the object label λ, F(R) is the feature measurements of the region R; e.g. average intensity, color, texture, aspect ratio, etc., and O means the output of the network for the input, F(R). The network W_λ operates as follows; If the region R labeled with λ exactly corresponds to that label, the network produces the lowest value; otherwise, it produces rather high values close to 1. Thus the domain knowledge about the object λ is implicitly stored in the connection weights of W_λ and W_λ plays a role as unary constraints for the object λ.

We construct and train the network independently for each label in Λ. In Figure 2 is shown the network for a certain object label. The network consists of input, output and one hidden layer. The number of input node is same as that of feature measurements and the number of output node is one.

F_i(R) denotes the ith feature measurements for the region R and f is the sigmoid function.

We trained the network so that it produce the lowest value close to 0 only if features are measured from the region whose exact object label is λ. Figure 3 shows the parameter values produced by the network, W_{road} given the feature measurements of the regions which are not included in the training examples. As you see in the figure, the network does not always produce the desired values for the "road" regions because there are so much variations on the input images due

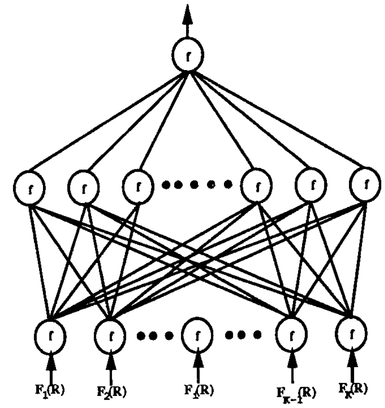

Figure 2: Error backpropagation network model for the single-node clique functions

to the lighting conditions, illuminations, etc. But we could get the correct labeling results for all the regions in experiments by exploiting the binary constraints based on the MRF model.

4 Image labeling by simulated annealing

After the region adjacency graph is formed, we assign randomly certain labels to every node on the graph and estimate the value of energy function through summatioin of all clique functions. The optimal interpretation that minimizes the energy function is then achieved by iterative operations until the energy function becomes stable. One iteration is defined as one visit to all the nodes in the graph, assigning new object label to the visited node and accepting the new label only if $e^{-\Delta E/T} > \mu$, where T is a temparature and μ is an random number between 0 and 1.

5 Experimental results

We applied the proposed method to the real-world images. In experiments, we used 6 object labels, 6 spectral and 2 geometric features for the unary constraints and one binary constraints as described in Section 2. Figure 4 shows the experimental results on a sample color image describing outdoor scene. Figure 4(b) shows the segmented regions generated by using the segmentation algorithm proposed by Suk et.al. [8], and Figure 4(d) shows the labeling results. Among segmented regions we ignored in labeling rather small and meaningless ones due to the complexity of the image, and which are remained as white regions in the labeling results. As depicted in Figure 4(e), from the beginning to about 20 iterations new labeling has been accepted although it resulted in a significant increase in energy function value. But after then, only a small increase in energy has been allowed and eventually, energy function reached the minimum and stabilized. Figure 4(f) shows the number of iterations to reach the minimum energy state in several experiments.

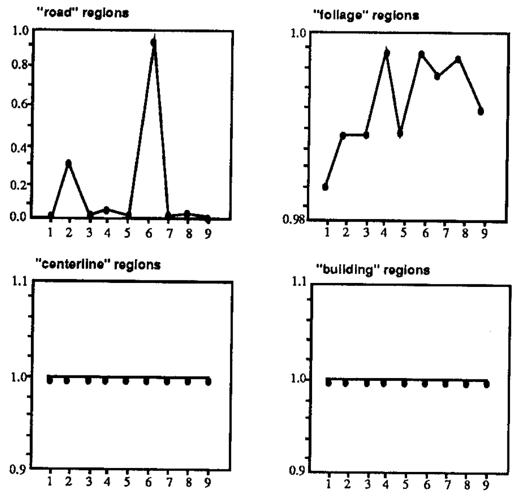

Figure 3: Parameter values produced by the network, W_{road} for the untrained regions:horizontal axes = number of sample regions, vertical axes = network output(α_ω)

6 Concluding remarks

In this paper, we have presented an efficient image labeling method based on the Markov Random Field. To endow the adaptability to the MRF-based image labeling, we have proposed a parameter estimation technique on error backpropagation. Through experiments on real natural scene images, we have confirmed the proposed method performs image labeling very effectively. In the future, in addition to the spatial adjacency, we would like to exploit more sturctured spatial relationships between regions to improve the performance.

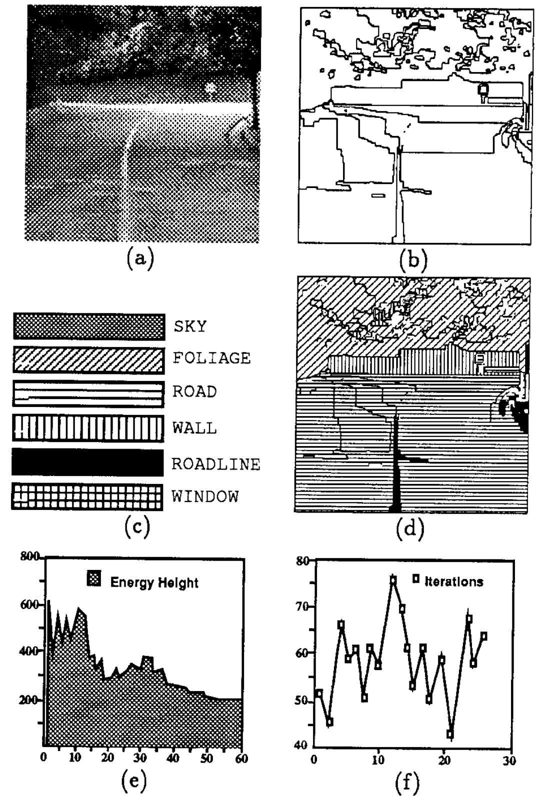

Figure 4: (a)real color image with 256x256 resolution (b)segmentation result (c) pattern key for real scene interpretation (d) labeling results (e)energy function values at each iteration (f)number of iterations to each correct labeling with minimum energy

References

[1] S.Geman and D.Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. PAMI-6, pp.721-741, Nov. 1984.

[2] H.Derin, H.Elliott and D.Geman, "Bayes smoothing algorithms for segmentation of binary images modeled by Markov Random Fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. PAMI-6, pp.707-720, Nov. 1984.

[3] S.Lakshmanan and H.Derin, "Simultaneous parameter estimation and segmentation of Gibbs Random Fields using simulated annealing," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. PAMI-11, pp.799-813, Aug. 1989.

[4] R.C.Dubes, A.K.Jain, S.G.Nadabar and C.C.Chen, "MRF model based algorithms for image segmentation," *Proc. Pattern Recognition*, Vol. 1, pp.808-814, 1990.

[5] J.W.Modestino and J.Zhang, "A Markov random field model based approach to image interpretation," *Proc. Comput. Vision Pattern Recognition*, pp.458-465, 1989.

[6] D.E.Rumelhart, G.E.Hinton and R.J.Williams, "Learning internal representation by error propagation," *Parallel Distributed Processing*, Vol. 1, MA:MIT Press, pp.318-362, 1986.

[7] S.Kirkpatrick, C.D.Gelatt and M.P.Vecchi, "Optimization by simulated annealing," *Science*, Vol. 220, pp.671-680, May 1983.

[8] M.S.Suk and S.M.Jung, "A new image segmentation technique based on Partition Mode Test," *Pattern Recognition*, Vol. 16, pp.469-480, 1983.