Stochastic Model-Based Image Segmentation Using Markov Random Fields and Multi-Layer Perceptrons

Jun Zhang[1]
Nelson Morgan[2]

TR-90-061

November , 1990

Abstract

Recently, there has been much interest in Markov random field (MRF) model-based techniques for image (texture) segmentation. MRF models are used to enforce reasonable physical constraints on segmented regions, such as the continuity of the regions, and have been shown to improve segmentation results. However, in these techniques, parametric probability models which do not have sufficient physical justifications are often used to model observed image data because they are computationally tractable. In this paper, we outline an MRF approach to image segmentation in which the probability distribution of observed image data is modeled by using a multi-layer perceptron (MLP) which can "learn" the distribution from training data. Furthermore, we propose a technique to achieve unsupervised image segmentation using this approach. We hope that this will improve the current MRF image segmentation techniques by providing a better model for observed image data.

[1]On leave from: Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI 53201.

[2]International Computer Science Institute, 1947 Center Street, Suite 600, Berkeley, CA 94704.

STOCHASTIC MODEL-BASED IMAGE SEGMENTATION USING MARKOV RANDOM FIELDS AND MULTI-LAYER PERCEPTRONS

JUN ZHANG[1] AND NELSON MORGAN
INTERNATIONAL COMPUTER SCIENCE INSTITUTE
1947 CENTER STREET, SUITE 600
BERKELEY, CA 94704

Abstract

Recently, there has been much interest in Markov random field (MRF) model-based techniques for image (texture) segmentation. MRF models are used to enforce reasonable physical constraints on segmented regions, such as the continuity of the regions, and have been shown to improve segmentation results. However, in these techniques, parametric probability models which do not have sufficient physical justifications are often used to model observed image data because they are computationally tractable. In this paper, we outline an MRF approach to image segmentation in which the probability distribution of observed image data is modeled by using a multi-layer perceptron (MLP) which can "learn" the distribution from training data. Furthermore, we propose a technique to achieve unsupervised image segmentation using this approach. We hope that this will improve the current MRF image segmentation techniques by providing a better model for observed image data.

[1]On leave from: Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI 53201.

1. Introduction:

Recently, there has been much interest in stochastic model-based image segmentation techniques. In such techniques, an image is separated into a set of disjoint regions that belong to a finite number of classes of different statistical properties. The image is usually modeled on two levels. On the first level, pixels that belong to the same class are modeled as a (conditional) random field. We will refer to the models on this level as the models for the image classes. On the second level, the spatial distribution of different image classes is also modeled by a random field known as the Markov random field (MRF). The MRF is used to enforce physical constraints, such as the continuity of the regions, to the resultant segmentation and has been shown to improve the quality of the segmentation [1]-[10].

Indeed, since the MRF model has been widely accepted as a model for the spatial distribution of regions, research is directed towards investigating the first level of image modeling, i.e., the models for the image classes. This involves trying out different texture models on, and learning such models from, real-world image data. The texture models used previously for image segmentation have been mostly parametric and largely Gaussian. While these parametric models (e.g., various Gaussian mixture models) may have simplified the structure of segmentation algorithms, they often lack "physical" justifications in the sense that they are not based on considerations of how images of different textures are formed.

As for learning, most previous techniques are supervised in that they assume the existence of training data. However, it is often desirable to have unsupervised learning, for example, when one wants a completely automated system (there is no operator intervention to select training data from some images, or regions of a single image). Bouman and Liu [11], Kelly et al. [12], Zhang [13], [15], and Shen [14] have proposed several unsupervised learning schemes. However, they all use parametric models for the image classes.

In this paper, we outline a non-parametric approach to modeling the observed image data. More specifically, we propose to use the multi-layer perceptron (MLP) for these models. MLPs, which are a form of "neural" network, can be trained to approximate many (or any, as claimed by Hecht-Nielsen [16]) multi-input multi-output functions. Therefore, when properly trained, they may provide a better model for the probability distribution of the image classes (or textures) than the somewhat arbitrarily selected parametric models. While MLPs ordinarily require supervised learning, we propose a "bootstrap" technique which may realize unsupervised learning.

The approach described in this paper is inspired by the work of Bourlard and Morgan [17] in speech recognition. In their application, the original speech signal is processed and features are extracted at time intervals known as frames. The likelihoods for these features can be calculated using an MLP, and the negative log likelihoods used as the local cost of a potential classification. These costs can then be accumulated to estimate the global cost of interpreting a continuous utterance as a particular sequence of phonemes. Bourlard and Morgan showed that when MLPs are used properly for estimating the joint likelihoods, the performance of the recognition system improves. In this paper, we attempt to use the MLP in an MRF approach to improve the performance of image segmentation procedures by generating better estimates of local likelihoods.

The organization of this paper is as follows. In Section 2, we formulate the image segmentation problem more precisely as a Bayesian estimation problem and describe an MRF approach to this problem. Then in Section 3, we outline how an MLP can be incorporated in the MRF approach and describe the computational requirement of such an approach. Finally, a summary is given in Section 4.

2. The MRF Approach:

Let $\mathbf{S}$ be a set of sites on a two-dimensional lattice on which the image is defined. We introduce the following notations:

$\mathbf{y} = \{y_i\}, i \in \mathbf{S}$ — the gray levels of image pixels;

$\mathbf{z} = \{z_i\}, i \in \mathbf{S}$ — the class indicator vectors for the pixels;

$p(\mathbf{z})$ — prior for $\mathbf{z}$;

$p(\mathbf{y}|\mathbf{z})$ — likelihood of $\mathbf{y}$.

We would like to make several remarks about these notations. First, a class indicator vector $z_i = (z_{i1}, z_{i2}, \ldots, z_{iK})$, where K is the number of image classes, is a binary vector with all but one component being zero. $z_{ik} = 1$ means y_i belongs to class k, $1 \leq k \leq K$. Secondly, the prior model $p(\mathbf{z})$ is used to enforce physical constraints on the segmentation, such as continuity (two adjacent pixels are likely to belong to the same class unless there is an edge between them), and is usually an MRF with a Gibbs distribution:

$$p(\mathbf{z}) = Z^{-1} \exp[-\beta U(\mathbf{z})], \tag{2.1}$$

where $U(\mathbf{z})$ is the energy function, Z is a normalization factor, and β is a positive control parameter. Thirdly, most previous work has adopted parametric models for $p(\mathbf{y}|\mathbf{z})$; we will not do so here. However, we do adopt an assumption made in most previous work, i.e.,

$$p(\mathbf{y}|\mathbf{z}) = \prod_i p(y_i|z_i). \tag{2.2}$$

As has been shown in a related work [15], this assumption can be relaxed.

Based on the notation introduced so far, image segmentation can be described as finding an estimate of $\mathbf{z}$ based on $\mathbf{y}$. Our approach is to find a maximum-a-posteriori (MAP) estimate, $\hat{\mathbf{z}}_{MAP}$, such that

$$\hat{\mathbf{z}}_{MAP} = \arg\max_{\mathbf{z}} p(\mathbf{z}|\mathbf{y}). \tag{2.3}$$

Using Bayes' conditional probability formula and noticing that the maximization in (2.3) is independent of $p(\mathbf{y})$, (2.3) is equivalent to

$$\hat{\mathbf{z}}_{MAP} = \arg\max_{\mathbf{z}} \{\log p(\mathbf{y}|\mathbf{z}) + \log p(\mathbf{z})\}. \tag{2.4}$$

Given the likelihood and the prior, the above maximization can be performed by simulated annealing [2], the ICM [5], and the mean field theory [20]. The first method, while potentially optimal, requires intensive computation. The second and the third are approximations to simulated annealing which require a moderate amount of computation. The prior is usually constructed before segmentation, while the likelihood has to be learned. Learning of the likelihood is the focus of the next section.

3. Learning by MLP:

Here, we consider two cases, supervised learning and unsupervised learning.

A. Supervised Learning:

This is the case in which training data is available. Let

$$\mathbf{T} = \{y_j, z_j\},\ j \in \mathbf{S}'$$

be a collection of training data, where y_j's are observed image data, z_j's are their known class indicator vectors, and $\mathbf{S}'$ is a collection of sites upon which the training data is defined. Then, according to a proposition in Bourlard and Morgan [17], a simple MLP, with a single input node (for y_j) m hidden nodes, and K output nodes, can learn the likelihood functional[2] $p(y|z)$. Here, K is the number of image (texture) classes, and m is a preset non negative number.

According to Bourlard and Morgan, the MLP works as follows. First, the MLP is trained by the back-propagation (BP) procedure on the training set $\mathbf{T}$, where y_j's are used as inputs and z_j's are used as target outputs. When the weights converge, the MLP will, for each input y_i (training or testing), produce approximately the probability $p(z_i|y_i)$ at its outputs. The likelihood functional $p(y_i|z_i)$ can then be obtained by dividing $p(z_i|y_i)$ by $p(z_i)$ (as estimated from the training data). Once the likelihood is learned, the segmentation can proceed as described in Section 2. Finally, when the likelihood model is more general than that of (2.2), for example, an MRF by itself, the structure of the MLP can be extended by including the neighboring pixels[3] of y_j as inputs.

B. Unsupervised Learning:

In this part, we propose an iterative (bootstrap) scheme for unsupervised learning of the likelihood. This scheme is inspired by similar parametric unsupervised learning techniques, such as the EM algorithm. Starting with an initial segmentation, denoted by $\hat{\mathbf{z}}^{(0)}$, each iteration in this scheme involves the following two steps:

Learning: learn the likelihood $p^{(n)}(y|z)$ by training the MLP on $\{y_i, \hat{z}_i^{(n-1)}\}$, $i \in S$, as described in A for supervised learning;

[2]Here we use $p(y|z)$ as a generic expression of the likelihood $p(y_i|z_i)$

[3]The neighboring pixels are, of course, determined by the neighborhood system of the MRF.

Segmentation: use the new likelihood, $p^{(n)}(y|z)$, to find a new segmentation $\hat{\mathbf{z}}^{(n)}$ by any of the maximization procedures described in Section 2.

Here, $n \geq 1$ represents the nth iteration. The iteration will stop if there is little change in the segmentation or enough iterations have been performed. To get the initial segmentation, one can simply generate a random segmentation or use any of the unsupervised parametric methods in [11]-[15].

C. Computational Requirements:

A preliminary analysis of the per-pixel calculations for a typical run of the above algorithm (using mean field calculations for the segmentation) shows that computation (as well as memory requirements) is dominated by the MLP training. To facilitate the MLP computation, either a special purpose computer such as the RAP [18],[19] or custom VLSI hardware can be used to implement the MLP. Most of the other computations for the segmentation stage (such as the estimation of the local mean field) can be accomplished by lookup of clique functions, exponentials, logarithms, etc. Assuming roughly 5 arithmetic operations per connection update, and something like 200 connections for a network to estimate probabilities for about 10 textures, and assuming about 10 iterations for the training, we would require about a billion operations for the training phase of a 100,000 pixel image. This would take a few seconds on the RAP machine or on a single special-purpose chip. While still too slow for real-time video applications, this is fast enough for research purposes, and would also permit more complex experimental networks. Initial experiments should be feasible for overnight runs on a general-purpose computer.

4. Summary:

In this paper, we have described an unsupervised MRF approach to image segmentation using an MLP. The MLP is used to learn the likelihood of the observed data through a bootstrap scheme. The potential advantage of this approach is that it may provide a better estimate of the likelihood than a (possibly) arbitrary parametric model. This approach is inspired by previous work on speech recognition where an MLP is used in an HMM approach and shown to have theoretical and practical advantages for recognition. Experiments have been planned to test out the approach described in this paper.

Acknowledgement:

We would like to thank Dr. Wolfgang Kuepper for his comments and suggestions, the National Science Foundation for an ROA award as an extension of grant MIP-8922354, and to the International Computer Science Institute for its continuing support of this work in connectionist applications.

References

[1] J. W. Modestino, R. W. Fries and A. L. Vickers, "Texture discrimination based upon an assumed stochastic texture model," IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-3, pp. 557-580, Sept., 1981.

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

[3] C. W. Therrien, T. F. Quatieri and D. E. Dudgeon, "Statistical model-based algorithms for image analysis", Proc. IEEE, Vol. 74, April, 1986.

[4] H. Derin and H. Elliot, "Modeling and segmentation of noisy and textured images using Gibbs random fields", IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-9, pp. 39-55, Jan., 1987.

[5] J. Besag, "On the statistical analysis of dirty pictures", J. Royal Stat. Soc. B., Vol. 48, pp. 259-302, 1986.

[6] J. Zhang, "Two-dimensional stochastic model-based image analysis", Ph.D. Thesis, Rensselaer Polytechnic Institute, Troy, New York, Aug., 1988.

[7] J. Marroquin, S. Mitter, and T. Poggio, "Computer vision", J. Amer. Stat. Association, Vol. 82, pp. 76-89, March, 1987.

[8] F. S. Cohen and D. B. Cooper, "Simple, parallel, hierarchical and relaxation algorithms for segmenting non-casual Markovian random field models", IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-9, pp. 195-219, March, 1987.

[9] J. F. Silverman and D. B. Cooper, "Bayesian clustering for unsupervised estimation of surface and texture models," IEEE Trans. Pattern Anal. Machine Intel., Vol. PAMI-10, pp. 482-495, July, 1988.

[10] T. N. Papas and N. S. Jayant, "An adaptive clustering algorithm for image segmentation," Proc. ICASSP'89, pp. 1667-1670. May, Glasgow, UK.

[11] C. Bouman and B. Liu, "Unsupervised estimation of image textures using an autoregressive model," Proc. ICASSP'90, pp. 2141-2144, 1990, Albuquerque, NM.

[12] P. A. Kelly, H. Derin, and K. D. Hart, "Adaptive segmentation of speckled images

using a hierarchical random field model," IEEE Trans. ASSP, Vol. 36, pp. 1628-1641, Oct., 1988.

[13] J. Zhang, "A ML approach for image segmentation," TR89-5, EE Dept., Univ. Wisc.-Milwaukee, under revision.

[14] Y. Shen, "Cluster validation with applications to image segmentation," M.S. Thesis, EE Dept., Univ. Wisc. -Milwaukee, Aug., 1990.

[15] J. Zhang, "The mean field theory in EM procedures for Markov random fields," ICSI TR, in revision.

[16] R. Hecht-Nielsen, "Theory of back propagation neural networks," Proc. IEEE-IJCNN89, Vol. I, pp. 593-605, Washington, D.C., 1989.

[17] H. Bourlard and N. Morgan, "Merging multilayer perceptrons and hidden Markov models: some experiments in continuous speech recognition," To be published in *Neural Networks: Advances and Applications*, Ed. E. Gelenbe, North-Holland.

[18] N. Morgan, J. Beck, E. Allman, and J. Beer, "RAP : A Ring Array Processor for Multilayer Perceptron Applications", Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, Albuquerque NM, 1990.

[19] N. Morgan, J. Beck, P. Kohn, J. Bilmes, E. Allman, and J. Beer, "The RAP : A Ring Array Processor for Layered Network Calculations", Proceedings Application Specific Array Processor Conference, pp. 296-308, 1990

[20] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, 1987.