Image Classification Based on a Multiresolution Two Dimensional Hidden Markov Model

Jia Li Robert M. Gray *

Abstract

This paper presents an image classification algorithm using a multiresolution two dimensional hidden Markov model (HMM). The multiresolution two dimensional hidden Markov model is an extension from the two dimensional hidden Markov model for image classification. A classifier estimates model parameters using the EM algorithm. Classification is then performed according to the maximum a posteriori probability criterion. The multiresolution model enables multiscale context information be incorporated into classification decisions. Suboptimal classification algorithms based on the model provide a progressive classification scheme which greatly speeds up classification with a single resolution HMM.

1 Introduction

In order to improve classification performance using context, a context-dependent classification algorithm is developed in [3] based on a two dimensional hidden Markov model (2-D HMM). Suppose an image is divided into blocks and feature vectors are evaluated for all the blocks, the basic assumption of a 2-D HMM is that the feature vectors are generated by a Markov source. At any block, the source exists in one of a set of states. Given the state of a block, its feature vector is assumed to follow a Gaussian distribution. The parameters of the Gaussian distribution vary with states. The probability distribution of the states of all the blocks is governed by a Markovian property, i.e., given the states of surrounding blocks, the probability of the source being in a state at a block depends upon the state of the source at the adjacent observations in both horizontal and vertical directions. Each state corresponds uniquely to a class. However, a class may contain several states. Details on the assumptions of a 2-D HMM may be found in [3].

We here extend the 2-D HMM to a multiresolution model. To classify an image, we evaluate feature vectors for image blocks in several resolutions. Feature vectors in a particular resolution are determined only by the image at that resolution. The multiresolution 2-D HMM assumes that the feature vectors across all the resolutions are generated by a multiresolution Markov source. As with the 2-D HMM, the source exists in a state at any block in any resolution. Given the state of a block in a particular resolution, the feature vector at the respective resolution is assumed to follow a Gaussian distribution. The parameters of the Gaussian distribution are determined by the state and the resolution. At any fixed resolution, as with the 2-D HMM, the probability of the source entering a particular state depends on the states of adjacent blocks in both horizontal and vertical directions. The transition probabilities, however, depend on the state of the block at the same spatial location in the previous resolution. The extension to a multiresolution model allows us to represent an image by features in several resolutions so that classification is made based upon more complete statistical properties of the image. Furthermore, the multiresolution model provides a hierarchical structure for progressive classification, which speeds up the 2-D HMM classification significantly.

In Section 2, we provide a mathematical formulation of the basic assumptions of multiresolution 2-D HMM. The algorithm is presented in Section 3. The results are given in Section 4.

2 Basic Assumptions of 2-D HMM

To classify an image, we first obtain several resolutions of the image. The original image corresponds to the highest resolution. We obtain lower resolutions by filtering out high frequency information. Wavelet transforms [2] naturally provide low resolution images by the low frequency band (the LL band). A sequence of images at several resolutions is shown in Fig. 2. As subsampling is applied for every reduced resolution, the image size decreases by a factor of

*The authors are with the Information Systems Laboratory, Department of Electrical engineering, Stanford University, CA 94305, U.S.A. This work was supported by the National Science Foundation under NSF Grant No. MIP-9706284. Email: jiali@isl.stanford.edu, rmgray@stanford.edu.

two in both directions. As is also shown in Fig. 2, the number of blocks in both rows and columns decreases to a half at each lower resolution. Thus a block in a lower resolution corresponds to a spatially more global region in the image. The four blocks at the same spatial location in the higher resolution are referred to as child blocks, and the block in the lower resolution is referred to as parent block correspondingly.

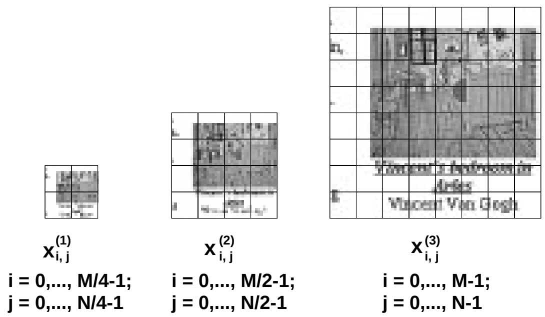

Figure 1: Multiple resolutions of an image.

We first review the basic assumptions of the single resolution 2-D HMM as presented in [3]. The 2-D HMM assumes that the feature vectors are generated by a Markov model which may change state once every block. Suppose there are M states, the state of block (i,j) is denoted by $s_{i,j}$. The feature vector of block (i,j) is $\mathbf{x}_{i,j}$ and the class is $c_{i,j}$. We use $P(\cdot)$ to represent the probability of an event. We denote $(i',j') < (i,j)$ if $i' < i$ or $i' = i, j' < j$; in which case we say that block (i',j') is before block (i,j). The first assumption we make is that $P(s_{i,j}|\text{context}) = a_{m,n,l}$, where context $= \{s_{i',j'}, \mathbf{x}_{i',j'}, (i',j') < (i,j)\}$, and $m = s_{i-1,j}$, $n = s_{i,j-1}$, and $l = s_{i,j}$. The second assumption is that for every state, the feature vectors follow a Gaussian distribution. Once the state of a block is known, the feature vector is conditionally independent of the other blocks. For a block with state s and feature vector $\mathbf{x}$, the distribution is $b_s(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n|\Sigma_s|}} e^{-\frac{1}{2}(\mathbf{x}-\mu_s)'\Sigma_s^{-1}(\mathbf{x}-\mu_s)}$, where Σ_s is the covariance matrix and μ_s is the mean vector. The parameters Σ_s and μ_s vary with states.

Suppose there are R resolutions, where $r = 1$ is the crudest resolution. The feature vectors at resolution r are $\mathbf{x}_{i,j}^{(r)}$, where (i,j) denotes block (i,j). Denote the collection of block indexes at resolution r by $\mathbb{N}^{(r)} = \{(i,j); 0 \le i < m \cdot 2^{r-1}, 0 \le j < n \cdot 2^{r-1}\}, r \in \mathbb{R}, \mathbb{R} = \{1, ..., R\}$.

The first assumption is a Markovian property across resolutions. The 'resolution' here plays a time-like role. Given the states and the features of the parent resolution, the states and the features of the current resolution are conditionally independent of the other previous resolutions, i.e.,

$$\begin{aligned} P(s_{i,j}^{(r)}, \mathbf{x}_{i,j}^{(r)}; r \in \mathbb{R}, (i,j) \in \mathbb{N}^{(r)}) &= P(s_{i,j}^{(1)}, \mathbf{x}_{i,j}^{(1)}; (i,j) \in \mathbb{N}^{(1)}) \cdot P(s_{i,j}^{(2)}, \mathbf{x}_{i,j}^{(2)}; (i,j) \in \mathbb{N}^{(2)} \,|\, s_{k,l}^{(1)}; (k,l) \in \mathbb{N}^{(1)}) \cdots \\ & \quad P(s_{i,j}^{(R)}, \mathbf{x}_{i,j}^{(R)}; (i,j) \in \mathbb{N}^{(R)} \,|\, s_{k,l}^{(R-1)}; (k,l) \in \mathbb{N}^{(R-1)}) \,. \end{aligned}$$

At the crudest resolution, $r = 1$, we assume that the feature vectors are simply generated by a single resolution 2-D HMM. At a higher resolution, we keep the assumption that given the state of a block, the feature vector follows a Gaussian distribution. We assume that given the state of the parent block, the states of the four child blocks are statistically independent of the states of the other blocks in the parent resolution. We also assume that given the states of parent blocks, the child blocks of different parent blocks are statistically independent. State transitions among the child blocks of a shared parent block are governed by the same Markovian property as we assumed for a single resolution 2-D HMM. The state transition probabilities, however, depend on the state of the parent block. To state these assumptions mathematically, denote the child blocks in resolution r of block (k,l) in resolution $r-1$ by $\mathbb{D}(k,l) = \{(2k,2l), (2k+1,2l), (2k,2l+1), (2k+1,2l+1)\}$. Then,

$$P(s_{i,j}^{(r)}; (i,j) \in \mathbb{N}^{(r)} \,|\, s_{k,l}^{(r-1)}; (k,l) \in \mathbb{N}^{(r-1)}) = \prod_{(k,l)\in\mathbb{N}^{(r-1)}} P(s_{i,j}^{(r)}; (i,j) \in \mathbb{D}(k,l) \,|\, s_{k,l}^{(r-1)}) \,.$$

$P(s_{i,j}^{(r)}; (i,j) \in \mathbb{D}(k,l) \,|\, s_{k,l}^{(r-1)})$ can be evaluated by the transition probabilities conditioned on $s_{k,l}^{(r-1)}$, $a_{m,n,l}(s_{k,l}^{(r-1)})$. We thus have a different set of transition probabilities $a_{m,n,l}$ for every possible state in the parent resolution. The influence of the previous resolutions is exerted hierarchically through the probability of the states.

As with the single resolution model, each state in every resolution is uniquely mapped to one class. Since a block in a lower resolution contain several blocks in a higher resolution, it may not be of a pure class. Thus, except for the highest resolution, there is an extra 'mixed' class besides the original classes.

3 The Algorithm

We estimate the parameters of the multiresolution model by the EM algorithm [1]. Due to computation complexity, we apply a suboptimal estimation algorithm, the Viterbi training algorithm [5]. The algorithm iteratively improves the estimation by searching for the combination of states with the maximum a posteriori probability given the observed feature vectors and the current model estimation. The states are then regarded as the true states to update the model estimation. Since our model is two dimensional, approximation to the Viterbi training is necessary to reduce computational complexity. Details on techniques for decreasing complexity are in [3]. Once the model is estimated, to classify an image, we search for the combination of states with the maximum a posteriori probability. The states are then mapped into classes to form the segmented image.

As states across resolutions are statistically dependent, determining the optimal set of states requires joint consideration of several resolutions. Suboptimal fast algorithms are developed by breaking the joint consideration and searching for the states in a layered fashion. States in the lowest resolution are determined by conditioning only on feature vectors in this resolution. The classifier searches for the states of the child blocks in the higher resolution, only if the class of the parent block is 'mixed'. Since one block in a lower resolution covers a larger region in the original image, making decisions in the lower resolution reduces computation. On the other hand, the existence of the 'mixed' class warns the classifier of ambiguous areas which need examination at higher resolutions. As a result, the degradation of classification due to the low resolution is avoided. The fast algorithm speeds up the single resolution HMM algorithm by a factor of $10 \sim 20$, while keeping competitive classification performance.

4 Results

We applied our algorithm to the segmentation of man-made and natural regions of aerial images. An example image and its hand-labeled classified image are shown on the left in Fig. 2. We use two resolutions in the classification. The lower resolution image is obtained by the LL frequency band of a Haar transform. For both resolutions, we divide images into 4×4 blocks and use DCT coefficients or averages over some of them as features. We used four images as training data to estimate the model. The number of states assumed for each class is 4. The classification error rate achieved is 12.12%. The classification error rate obtained by the single resolution HMM in [3] is 14.68%. The classified images are compared in Fig. 2. Using a fast algorithm, we achieve an error rate of 13.01%. The error rate obtained by Bayes VQ [4] is about 21.5%.

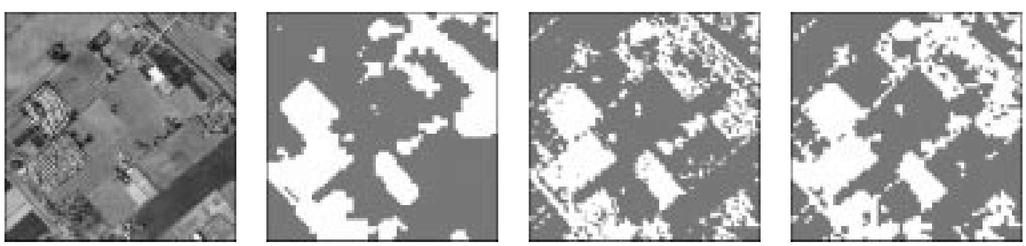

Figure 2: Left: Original image and its hand-labeled classes. Middle: Single resolution HMM, Right: Multiresolution HMM (2 resolutions), White: man-made, Gray: natural.

References

[1] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Statist. Soc.*, 1997.

[2] I. Daubechies, *Ten Lectures on Wavelets*, Capital City Press, 1992.

[3] J. Li, A. Najmi and R. M. Gray, "Image Classification by a Two Dimensional Hidden Markov Model," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, Arizona, March, 1999.

[4] K. O. Perlmutter, S. M. Perlmutter, R. M. Gray, R. A. Olshen, and K. L. Oehler, "Bayes Risk Weighted Vector Quantization with Posterior Estimation for Image Compression and Classification," *IEEE Transactions on Image Processing*, vol. 5, no. 2, pp. 347-60, February 1996.

[5] S. Young, J. Jansen, J. Odell, et al., *HTK - Hidden Markov Model Toolkit*, Cambridge University, 1995.