A GENERAL TWO-DIMENSIONAL HIDDEN MARKOV MODEL AND ITS APPLICATION IN IMAGE CLASSIFICATION

Xiang Ma, Dan Schonfeld and Ashfaq Khokhar

Department of Electrical and Computer Engineering, University of Illinois at Chicago
851 S Morgan St., Chicago, IL 60607, U.S.A.
{mxiang, ds, ashfaq}@ece.uic.edu

ABSTRACT

In this paper, we propose a general two-dimensional hidden Markov model (2D-HMM), where dependency of the state transition probability on any state is allowed as long as causality is preserved. The proposed 2D-HMM model can capture, for example, dependency among diagonal states, which can be critical in many image processing applications. A new Expectation-Maximization (EM) algorithm suitable for estimation of the new model is derived, where a novel General Forward-Backward (GFB) algorithm is proposed for recursive estimation of the model parameters. A new conditional-independent subset-state sequence structure decomposition is proposed for the 2D Viterbi algorithm. The new model can be applied to many areas such as trajectory classification and image segmentation. Application to aerial image segmentation shows the superiority of our model compared to the existing 2D-HMM model.

Index Terms— Image classification, Hidden Markov models, Image segmentation.

1. INTRODUCTION

In most block-based image classification algorithms, feature vectors are generated for each image block. Classification decisions are made independently for each block based on feature information. The performance of such algorithms is limited since context information between blocks is lost. J. Li et al. [1] proposed a two-dimensional hidden Markov model for image classification, where state transition probability for each block is conditioned on the states of nearest neighboring blocks from horizontal and vertical directions. However, the context information that a block depends on may arise from any direction and from any of its neighbors as shown in Fig. 1(a). Thus, the existing two-dimensional model will only capture partial context information. Generalization of the hidden Markov model (HMM) framework to represent state dependencies from neighbors in all directions is unsolvable since such a model will be non-causal. In this paper, a general two-dimensional hidden Markov model (2D-HMM) is proposed. The context information in the proposed model is restricted to neighbors which ensure the causality of the model as shown in Fig. 1(a). For simplicity, the presentation in this paper will focus primarily on a special case of the general 2D-HMM model, where dependencies arise from adjacent diagonal, horizontal and vertical neighbors for each block as shown in Fig. 1(b). The main challenge is that the existing methods can not be used to solve the proposed model. We must therefore derive new algorithms for training, estimation and classification, i.e. (1) Newly derived Expectation-Maximization (EM) algorithm; (2) General Forward-Backward (GFB) algorithm; (3) 2D Viterbi algorithm suitable for the structure of the proposed model.

The rest of paper is organized as follows: In Section 2, we provide a mathematical formulation of the proposed model. Section 3 will discuss proposed algorithms for 2D-HMM training and classification. Experimental results of the proposed model for the segmentation of aerial images are demonstrated in Section 4. We finally conclude in Section 5.

This work is funded in part of funding from NSF IIS-0534438.

2. PROPOSED TWO-DIMENSIONAL HIDDEN MARKOV MODEL (2D-HMM)

The proposed model follows the assumptions below:

Assumption 1 *The transition probability of state $s(i,j)$ in the model depends on its adjacent neighboring states in vertical, horizontal and diagonal directions.*

Suppose there are M states $\{1,2,...,M\}$, and for each block (i,j), $i=\{1,2,...,I\}$; $j=\{1,2,...,J\}$, where I and J are the numbers of row and column blocks in the original image, the feature vector is $o(i,j)$, the corresponding hidden state is $s(i,j)$, and the class of the block is $c(i,j)$. We define the transition probability of state $s(i,j)$ and it depends on its adjacent neighboring states in vertical, horizontal and diagonal directions, as shown in Figure 1(b), and is stated as follows:

$$P\{s(i,j)=l|s(i-1,j)=m, s(i-1,j-1)=n, s(i,j-1)=k\} = a_{m,n,k,l}. \quad (1)$$

where $m,n,k,l \in \{1,2,...,M\}$ are actual values of the state.

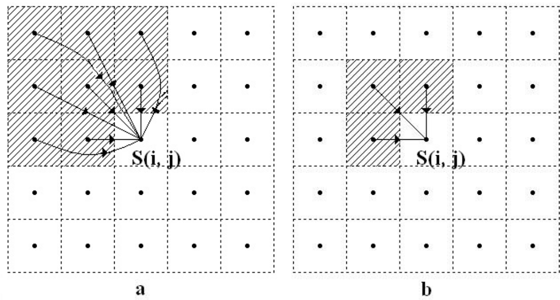

Fig. 1. Example of state transition diagrams of 2D-HMM: (a) general case (b) one simple case (in both cases, only part is shown). Dashed square: image block, Dot: state, Shaded blocks: neighbors of block $s(i,j)$.

The above assumption guarantees that for each state, its adjacent neighbors in horizontal, vertical and diagonal directions will be "before" it in the sense of spatial localization, thus ensuring the causality and Markovian property of the model.

Assumption 2 *The feature vector for each image block follows a Gaussian Mixture distribution, given its corresponding state, and it is independent of other feature vectors and their corresponding states.*

Since any state with an M-component Gaussian mixture can be split into M substates with single Gaussian distributions[1], we define the probability density function of feature vector $o(i,j)$, given its corresponding hidden state $s(i,j)=m$, as

$$b_m(o(i,j)) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_m|^{\frac{1}{2}}} e^{-\frac{1}{2}(o(i,j)-\mu_m)^T\Sigma_m^{-1}(o(i,j)-\mu_m)} \tag{2}$$

where d is the dimensionality of feature vector, μ_m and Σ_m are the mean vector and covariance matrix of Gaussian distribution corresponding to state m. The proposed model can be seen in Figure 2(a), which satisfies the above two assumptions.

3. 2D-HMM TRAINING AND CLASSIFICATION

3.1. Expectation-Maximization (EM) algorithm

We propose a newly derived Expectation-Maximization (EM) algorithm suitable for the estimation of parameters of proposed 2D-HMM model, which is analog but different than the EM algorithm for 1D HMM [2][3]. Define the observed feature vector set $O=\{o(i,j), i=1,2,...I; j=1,2,...J\}$ and corresponding hidden state set $S=\{s(i,j), i=1,2,...I; j=1,2,...J\}$. The model parameters are defined as a set $\Theta=\{\Pi,\mathbf{A},\mathbf{B}\}$, where $\Pi=\{\pi_m\}$ is the set of initial probabilities of states; $\mathbf{A}=\{a_{m,n,k,l}\}$ is the set of state transition probabilities, $(m,n,k,l\in\{1,2,...,M\})$; and **B** is the set of probability density functions (PDFs) of the observed feature vectors given corresponding states.

Define $F^{(p)}_{m,n,k,l}(i,j)$ as the probability of state corresponding to observation $o(i-1,j)$ is state m, state corresponding to observation $o(i-1,j-1)$ is state n, state corresponding to observation $o(i,j-1)$ is state k and state corresponding to observation $o(i,j)$ is state l, given the observations and model parameters,

$$F^{(p)}_{m,n,k,l}(i,j) = P\Big(m=s(i-1,j), n=s(i-1,j-1), k=s(i,j-1), l=s(i,j)|O,\Theta^{(p)}\Big), \tag{3}$$

and define $G^{(p)}_m(i,j)$ as the probability of the state corresponding to observation $o(i,j)$ is state m, then

$$G^{(p)}_m(i,j) = P(s(i,j)=m|O,\Theta^{(p)}). \tag{4}$$

We can get the iterative updating formulas of parameters of the proposed model,

$$\pi^{(p+1)}_m = P(G^{(p)}_m(1,1)|O,\Theta^{(p)}). \tag{5}$$

$$a^{(p+1)}_{m,n,k,l} = \frac{\sum_i^I\sum_j^J F^{(p)}_{m,n,k,l}(i,j)}{\sum_{l=1}^M\sum_i^I\sum_j^J F^{(p)}_{m,n,k,l}(i,j)}. \tag{6}$$

$$\mu^{(p+1)}_m = \frac{\sum_i^I\sum_j^J G^{(p)}_m(i,j)o(i,j)}{\sum_i^I\sum_j^J G^{(p)}_m(i,j)}. \tag{7}$$

$$\Sigma^{(p+1)}_m = \frac{\sum_i^I\sum_j^J G^{(p)}_m(i,j)(o(i,j)-\mu^{(p+1)}_m)(o(i,j)-\mu^{(p+1)}_m)^T}{\sum_i^I\sum_j^J G^{(p)}_m(i,j)}. \tag{8}$$

In eqns. (3)-(8), p is the iteration step number. $F^{(p)}_{m,n,k,l}(i,j)$, $G^{(p)}_m(i,j)$ are unknown in the above formulas, next we propose a General Forward-Backward (GFB) algorithm for the estimation of them.

3.2. General Forward-Backward (GFB) algorithm

Forward-Backward algorithm was firstly proposed by Baum et al. [4] for 1D Hidden Markov Model. Later, Jia Li et al. [1] proposed a similar Forward-Backward algorithm for their model. Here, we would like to generalize the Forward-Backward algorithm in [1][4] so that it can be applied to any HMM system, the proposed algorithm is called General Forward-Backward (GFB) algorithm.

For any HMM system, if its state sequence satisfy the following property, then GFB algorithm can be applied to it:

Property 1 *The probability of all-state sequence S can be decomposed as products of probabilities of conditional-independent subset-state sequences $U_0, U_1, ...$.*

$$P(S) = P(U_0)P(U_1/U_0)...P(U_i/U_{i-1})... \tag{9}$$

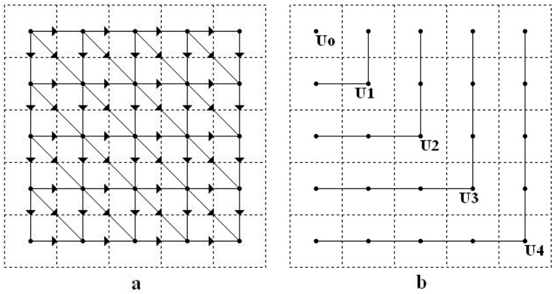

Fig. 2. (a) State transition diagram of proposed 2D-HMM and (b) its decomposed subset-state sequences.

where $U_0, U_1, ..., U_i$...are subsets of all-state sequence in the HMM system, we call them *subset-state sequences*. Define the observation sequence corresponding to each subset-state sequence U_i as O_i. Subset-state sequences for our model are shown in Figure 2(b), which is similar to [5]. Please note that if the dependencies are not defined on adjacent nearest neighbors, U_i are no longer guaranteed to be non-overlapping. The new structure enables us to use General Forward-Backward (GFB) algorithm to estimate the model parameters.

3.2.1. Forward and Backward Probability

Definition 1 *The forward probability $\alpha_{U_u}(u), u = 1, 2, ...$ is the probability of observing the observation sequence $O_v(v \leq u)$ corresponding to subset-state sequence $U_v(v \leq u)$ and having state sequence for u-th product component in the decomposing formula as U_u, given model parameters Θ.*

$$\alpha_{U_u}(u) = P\{S(u) = U_u, O_v, v \leq u|\Theta\} \tag{10}$$

The recursive updating formula of forward probability is

$$\alpha_{U_u}(u) = [\sum_{u-1} \alpha_{U_{u-1}}(u-1)P\{U_u|U_{u-1}, \Theta\}]P\{O_u|U_u, \Theta\}. \tag{11}$$

for $u > 1$.

Definition 2 *The backward probability $\beta_{U_u}(u), u = 1, 2, ...$ is the probability of observing the observation sequence O_v $(v > u)$ corresponding to subset-state sequence $U_v(v > u)$, given state sequence for u-th product component as U_u and model parameters Θ.*

$$\beta_{U_u}(u) = P(O_v, v > u|S(u) = U_u, \Theta). \tag{12}$$

We also derive the recursive updating formula of backward probability as follows:

$$\beta_{U_u}(u) = \sum_{u+1} P(U_{u+1}|U_u, \Theta)P(O_{u+1}|U_{u+1}, \Theta)\beta_{U_{u+1}}(u+1). \tag{13}$$

The estimation formulas of $F_{m,n,k,l}(i, j)$, $G_m(i, j)$ are :

$$G_m(i, j) = \frac{\alpha_{U_u}(u)\beta_{U_u}(u)}{\sum_{u:U_u(i,j)=m} \alpha_{U_u}(u)\beta_{U_u}(u)}. \tag{14}$$

$$F_{m,n,k,l}(i, j) = \frac{\alpha_{U_{u-1}}(u-1)P(U_u|U_{u-1}, \Theta)P(O_u|U_u, \Theta)\beta_{U_u}(u)}{\sum_u \sum_{u-1}[\alpha_{U_{u-1}}(u-1)P(U_u|U_{u-1}, \Theta)P(O_u|U_u, \Theta)\beta_{U_u}(u)]}. \tag{15}$$

3.3. 2D Viterbi algorithm

For classification, we employ a two-dimensional Viterbi algorithm [5] to search for the best combination of states with maximum a posteriori probability and map each block to a class. This process is equivalent to search for the state of each block using an extension of the variable-state Viterbi algorithm presented in [1], based on the new structure in Fig. 2(b).

If we search for all the combinations of states, suppose the number of states in each subset-state sequence U_u is $w(u)$, then the number of possible sequences of states at every position will be $M^{w(u)}$, which is computationally infeasible. To reduce the computational complexity, we only use N sequences of states with highest likelihoods out of the $M^{w(u)}$ possible states.

3.4. Summary of the proposed algorithms

-*Training*:

1. Assign initial values to π_m, $a_{m,n,k,l}$, μ_m and Σ_m.
2. Update the forward and backward probabilities according to equations (11) and (13) using proposed GFB algorithm.
3. Update $F_{m,n,k,l}(i, j)$, $G_m(i, j)$ according to equations (14)-(15).
4. Update $\{\pi_m, a_{m,n,k,l}, \mu_m, \Sigma_m\}$ according to equations (5)-(8) using proposed EM algorithm.
5. Back to step 2, stop until changes of parameters are below pre-set thresholds.

-*Classification*:

Use the proposed two-dimensional Viterbi algorithm to search for the best combination of states with maximum a posteriori probability.

4. EXPERIMENTAL RESULTS: IMAGE CLASSIFICATION

In this section, we compare our general 2D-HMM model with the model presented in [1] for the classification and segmentation of man-made and natural regions of aerial images. The images used are 6 aerial images of the San Francisco Bay area provided by TRW (formerly ESL, Inc.). One of the six images used is shown in Fig. 3(a) and its hand-labeled truth

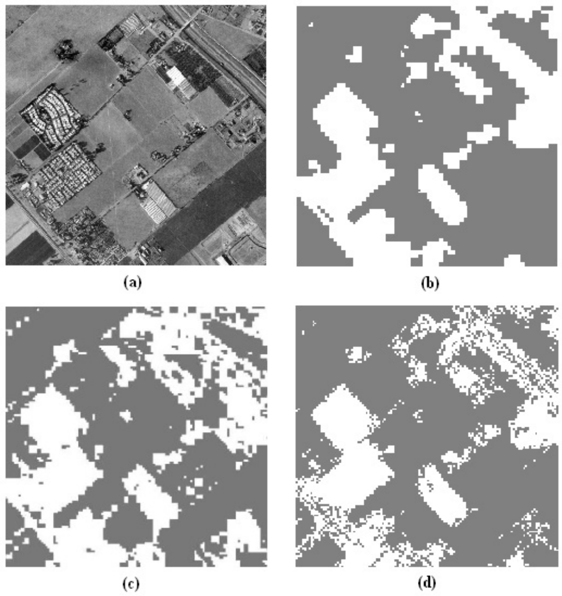

Fig. 3. Comparison of the classification results of the proposed general 2D-HMM model and the model presented in [1]: (a) an original aerial image; (b) hand-labeled truth image; (c) classification results using the model presented in [1]—error rate 13.39%; and, (d) classification results using the proposed general model—error rate 11.67%. (White: man-made regions, Gray: natural regions)

image is depicted in Fig. 3(b). The images are divided into non-overlapping blocks, and feature vectors for each block are extracted. The feature vector consists of nine features, of which 6 are intra-block features, as defined in [1], and 3 are inter-block features defined as the differences of average intensities of block (i, j) with its vertical, horizontal and diagonal block. Let the average intensity of block (i, j) be $\overline{I}(i, j)$, then the 3 features are $f_7 = \overline{I}(i, j) - \overline{I}(i-1, j)$; $f_8 = \overline{I}(i, j) - \overline{I}(i-1, j-1)$; and $f_9 = \overline{I}(i, j) - \overline{I}(i-1, j)$.

We first train our model using training images, and estimate the model parameters based on the training feature vectors and their corresponding truth set of classes. We then perform image classification for a test image using the trained model. Feature vectors are generated for each block in the test image in the same way as in training. For testing the model, six-fold cross-validation is used. For each test, one image is used as a test image, and the other 5 serve as training images. 2D-HMM models with different number of states and different block sizes are evaluated. We found that the model with 6 states for the natural class and 8 states for the man-made class yields the best result. The classification results are shown in Table 1. We can see that for all of the different block sizes, the proposed 2D-HMM model outperforms the existing model. Comparison of these models for one of the classified images is shown in Figs. 3(c) and 3(d).

Table 1. Average Classification Error Rate VS. Blocksize

Algorithm	Blocksize=4	Blocksize=8	Blocksize=16
Existing model	0.1874	0.2420	0.2921
Proposed model	**0.1536**	**0.2041**	**0.2596**

5. CONCLUSIONS

We propose a general two-dimensional hidden Markov model. This model allows state dependency in diagonal directions. Our approach to 2D-HMM model can be extended beyond the diagonal direction to any dependency which preserves causality. A new Expectation-Maximization algorithm, a General Forward-Backward algorithm and a new structure of the 2D Viterbi algorithm are proposed. The application to aerial image classification shows the superior performance of our 2D-HMM model than the existing model. The proposed model can be extended further to a more general non-causal, multidimensional hidden Markov model by decomposing it into multiple distributed causal HMMs. Preliminary results have shown that this method will preserve much more information and yield superior results. This approach will be explored in our future work.

6. ACKNOWLEDGEMENT

The authors would like to thank Prof. Jia Li from Penn. State Univ. for very helpful discussions and for providing the test images.

7. REFERENCES

[1] J. Li, A. Najmi, and R. M. Gray, “Image classification by a two-dimensional hidden markov model,” *IEEE Trans. on Signal Processing*, vol. 48, pp. 517–533, 2000.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *Journal of the Royal Statistical Society: Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[3] J. A. Bilmes, “A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models,” *Dept. of EECS, U. C. Berkeley, Berkeley, CA 94704, TR-97-021*, 1998.

[4] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, “A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains,” *Ann. Math. Stat.*, vol. 1, pp. 164–171, 1970.

[5] D. Schonfeld and N. Bouaynaya, “A new method for multidimensional optimization and its application in image and video processing,” *IEEE Signal Processing Letters*, vol. 13, pp. 485–488, 2006.