DISTRIBUTED MULTI-DIMENSIONAL HIDDEN MARKOV MODELS FOR IMAGE AND TRAJECTORY-BASED VIDEO CLASSIFICATIONS

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.

ABSTRACT

In this paper, we propose a novel multi-dimensional distributed hidden Markov model (DHMM) framework. We first extend the theory of 2D hidden Markov models (HMMs) to arbitrary causal multi-dimensional HMMs and provide the classification and training algorithms for this model. The proposed extension of causal multi-dimensional HMMs allows state transitions in arbitrary causal directions and neighbors. We subsequently generalize this framework further to non-causal models by distributing the non-causal models into multiple causal multi-dimensional HMMs. The proposed training and classification process consists of the extension of three fundamental algorithms to multi-dimensional causal systems, i.e. (1) Expectation-Maximization (EM) algorithm; (2) General Forward-Backward (GFB) algorithm; and (3) Viterbi algorithm. Simulation results performed using real-world images and videos demonstrate the superior performance, higher accuracy rate and promising applicability of the proposed DHMM framework.

Index Terms— Hidden Markov Models, Image Classification, Trajectory Classification.

1. INTRODUCTION

Hidden Markov Models (HMMs) have received tremendous attention in recent years due to its wide applicability in diverse areas such as speech recognition and trajectory classification. Most of the previous research has focused on the classical one-dimensional HMM developed in the 1960s by Baum et al [1], where the states of the system form a single one-dimensional Markov chain. However, the one-dimensional structure of this model limits its applicability to more complex data elements such as images and videos.

In this paper, we propose a novel multi-dimensional distributed hidden Markov model (DHMM) framework. We first provide a solution for non-causal, multi-dimensional HMMs by distributing the non-causal model into multiple distributed causal HMMs. We approximate the simultaneous solution of multiple distributed HMMs on a sequential processor by an alternate updating scheme. Subsequently we extend the training and classification algorithms presented in [2] to a general causal model. The proposed DHMM model can be applied to many problems in pattern analysis and classification.

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

2. DISTRIBUTED MULTI-DIMENSIONAL HIDDEN MARKOV MODEL: THEORY

We propose a novel solution to arbitrary non-causal multi-dimensional hidden Markov model, by distributing it into multiple causal distributed hidden Markov models and process them simultaneously.

For an arbitrary non-causal two-dimensional hidden Markov model which has N^2 state nodes lying on the two-dimensional state transitional diagram, if every dimension of the model is non-causal, we can solve the model by allocating N^2 processors, each for one node, and if the N^2 processors can be perfectly synchronized and dead-lock of concurrent state dependencies can be successfully solved, we can estimate the parameters of the non-causal model by setting all N^2 processors working simultaneously in perfect synchrony. However, this is usually impractical in reality. We propose to distribute the non-causal model to N^2 distributed causal models, by specifically focusing on the state dependencies of each node one at a time, while ignoring other nodes. Similarly, for arbitrary M-dimensional hidden Markov models, we can distributing the non-causal model to N^M distributed causal HMMs, by specifically focusing on the state dependencies of each node one at a time, while ignoring other nodes.

Fig. 1 depicts state dependencies diagrams of one non-causal two-dimensional model (Fig. 1(a)) and its decomposed two causal models (Fig. 1(b) and 1(c)). Directions of arrow show state dependencies, e.g. state node A points to B means that B depends on A. We refer to the distributed causal hidden Markov models as DHMMs.

Please note that in the distributing procedure, the state dependency information is not lost but considered. Further more, since each distributed submodel preserved the correlation between neighboring state nodes, the proposed framework is not a simple collection of uncorrelated causal models but an accurate representation of the original model. To accurately estimate the state transition probabilities of the non-causal model, all of the distributed causal two-dimensional models must be processed simultaneously in perfect synchrony. However, in reality, it is impossible for the whole

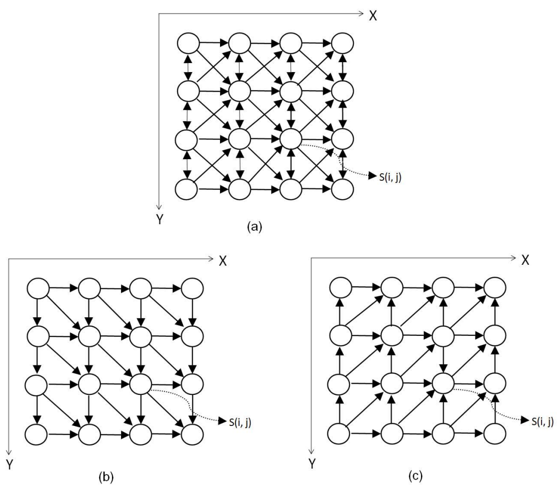

Fig. 1. DHMM example: (a) Non-causal 2D HMM. (b) Distributed Causal 2D HMM 1. (c) Distributed Causal 2D HMM 2.

system to be exactly synchronous. We pre-define an updating sequence of all state parameters of the distributed in order to achieve optimal approximation of the true model parameters.

3. DHMM TRAINING AND CLASSIFICATION

Define the observed feature vector set $O = \{(i,j), i = 1,2,...,I; j = 1,2,...,J\}$ and corresponding hidden state set $S = \{s(i,j), i = 1,2,...,I; j = 1,2,...,J\}$, and assume each state will take Q possible values. The model parameters are defined as a set $\Theta = \{\Pi, \mathbf{A}, \mathbf{B}\}$, where Π is the set of initial probabilities of states $\Pi = \{s(1,1)\}$; $\mathbf{A}$ is the set of state transition probabilities $A = \{a_{m,n,k,l}\}$, and

$$a_{m,n,k,l} = Pr(s(i,j) = l | s(i',j) = k; s(i,j-1) = m, s(i',j-1) = n) \tag{1}$$

and B is the set of probability density functions (PDFs) of the observed feature vectors given corresponding states, assume B is a set of Gaussian distribution with means μ_m and variances $\Sigma_{m,n}$, where $m,n,k,l = 1,...,Q; i \neq i'; i = 1,...,I; j = 1,...,J$.

3.1. Expectation-Maximization (EM) algorithm

We propose a newly derived Expectation-Maximization (EM) algorithm suitable for the estimation of parameters of the proposed model, which is an extention of the classical EM algorithm [3] to higher dimensions.

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, i.e.

$$F^{(p)}_{m,n,k,l}(i,j) = P\bigg(m = s(i-1,j), n = s(i-1,j-1), k = s(i,j-1), l = s(i,j) | O, \Theta^{(p)}\bigg), \tag{2}$$

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

$$G^{(p)}_m(i,j) = P(s(i,j) = m | O, \Theta^{(p)}). \tag{3}$$

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

$$\pi^{(p+1)}_m = P(G^{(p)}_m(1,1) | O, \Theta^{(p)}). \tag{4}$$

$$a^{(p+1)}_{m,n,k,l} = \frac{\sum_i^I \sum_j^J F^{(p)}_{m,n,k,l}(i,j)}{\sum_{l=1}^M \sum_i^I \sum_j^J F^{(p)}_{m,n,k,l}(i,j)}. \tag{5}$$

$$\mu^{(p+1)}_m = \frac{\sum_i^I \sum_j^J G^{(p)}_m(i,j) o(i,j)}{\sum_i^I \sum_j^J G^{(p)}_m(i,j)}. \tag{6}$$

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

In eqns. (4)-(7), 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

We generalize the Forward-Backward algorithm in [1] [2] so that it can be applied to any HMM system, the proposed algorithm is called General Forward-Backward (GFB) algorithm.

We decompose the state sequence S of the proposed model as follows:

$$P(S) = P(U_0)P(U_1/U_0)...P(U_i/U_{i-1})... \tag{8}$$

where $U_0, U_1, ..., U_i$...are subsets of all-state sequence S , 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 Fig. 2(b), which is similar to [4]. 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.*

$$\alpha_{U_u}(u) = P\{S(u) = U_u, O_v, v \leq u | \Theta\} \tag{9}$$

The recursive updating formula of forward probability is

$$\alpha_{U_u}(u) = [\sum_{u-1} \alpha_{U_{u-1}}(u-1) P\{U_u | U_{u-1}, \Theta\}] P\{O_u | U_u, \Theta\}. \tag{10}$$

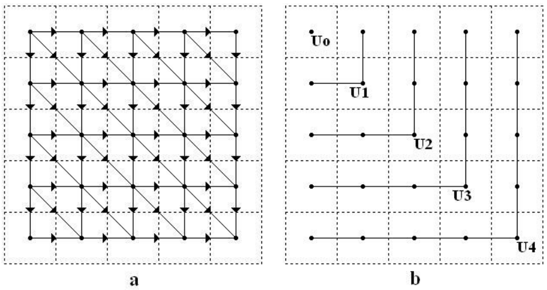

Fig. 2. (a) State transition diagram of proposed 2D-HMM and (b) its decomposed subset-state sequences.

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.

$$\beta_{U_u}(u) = P(O_v, v > u|S(u) = U_u, \Theta). \tag{11}$$

We also derive the recursive updating formula of backward probability as follows ($u > 1$):

$$\beta_{U_u}(u) = \sum_{u+1} P(U_{u+1}|U_u, \Theta)P(O_{u+1}|U_{u+1}, \Theta)\beta_{U_{u+1}}(u+1). \tag{12}$$

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

$$G_m(i, j) = \frac{\alpha_{U_u}(u)\beta_{U_u}(u)}{\sum_{u:U_u(i,j)=m} \alpha_{U_u}(u)\beta_{U_u}(u)}. \tag{13}$$

$$F_{m,n,k,l}(i, j) =$$
$$\frac{\alpha_{U_{u-1}}(u-1)P(U_u|U_{u-1}, \Theta)P(O_u|U_u, \Theta)\beta_{U_u}(u)}{\sum_u \sum_{u-1}[\alpha_{U_{u-1}}(u-1)P(U_u|U_{u-1}, \Theta)P(O_u|U_u, \Theta)\beta_{U_u}(u)]}. \tag{14}$$

3.3. 2D Viterbi algorithm

For classification, we employ a two-dimensional Viterbi algorithm [4] 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 [2], based on the new structure in Fig. 2(b).

3.4. Summary of DHMM Training and Classification Algorithms

-Training:

1. Assign initial values to $\{\pi_m, a_{m,n,k,l}, \mu_m, \Sigma_m\}$.
2. Update the forward and backward probabilities according to eqns. (10) and (12) using proposed GFB algorithm, calculate old $logP(O|\Theta_0)$.
3. Update $F_{m,n,k,l}(i, j)$, $G_m(i, j)$ according to eqns. (13)(14).
4. Update π_m, $a_{m,n,k,l}$, μ_m and Σ_m according to eqns. (4)-(7) using the proposed EM algorithm.
5. Back to step 2,Calculate new $logP(O|\Theta)$, stop if $logP(O|\Theta)$-$logP(O|\Theta_0)$ is below pre-set threshold.

-Classification: Use a two-dimensional Viterbi algorithm to search for the best combination of states with maximum a posteriori (MAP) probability.

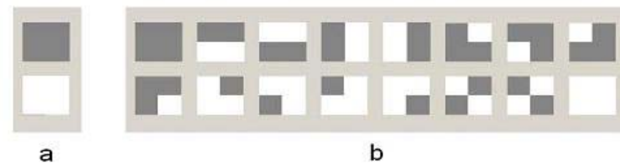

Fig. 3. Image block patterns: (a) 2 patterns in [2] (b) proposed 16 basic image block patterns (White: man-made regions, Gray: natural regions).

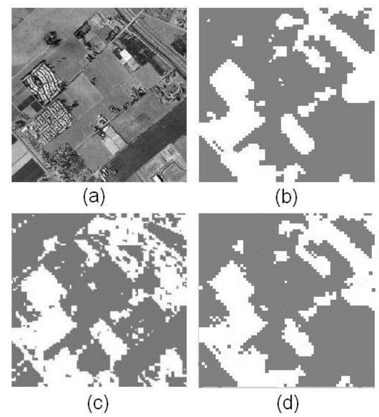

Fig. 4. Comparison of image classification results: (a) an original aerial image; (b) hand-labeled truth image; (c) classification results using the model presented in [2]—error rate 13.39%; and, (d) classification results using the proposed general model with 16 basis image block patterns—error rate **8.25%**. (White: man-made regions, Gray: natural regions)

4. APPLICATION I: DHMM-BASED REAL-WORLD IMAGE CLASSIFICATION

In this section, we test our DHMM model for the segmentation of man-made and natural regions of 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. 4(a) and its hand-labeled truth image is depicted in Fig. 4(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 six are intra-block features, as defined in [2], and three are inter-block features defined as the differences of average intensities of block (i, j) with its vertical, horizontal and diagonal block.

In previous work [2], image classification decisions are made for each block, either man-made region or natural region, based its corresponding hidden states. However, image blocks are not necessarily total man-made or natural region. In reality, most image blocks are mixture of man-made and natural regions. Based on this observation, we propose to define 16 basic image block patterns that cover all possible variabilities of image blocks, depicted in Fig 3(b). An image block can be either totally man-made, or natural, or mixture of man-made and natural regions. Each pattern has several corresponding hidden states, which enriches the variability

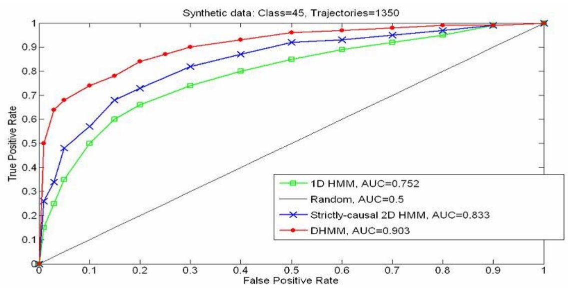

Fig. 5. ROC curve of DHMM, Strictly Causal 2D HMM and 1D HMM for Synthetic data

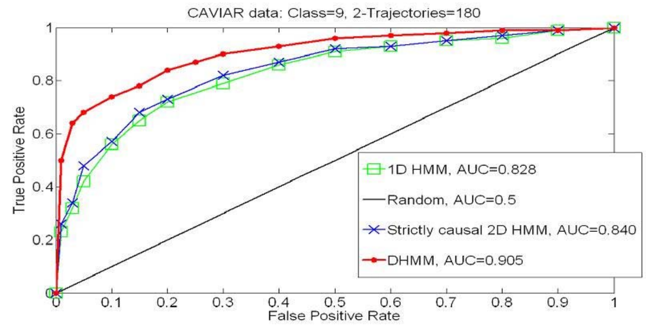

Fig. 6. ROC curve of DHMM, Strictly Causal 2D HMM and 1D HMM for CAVIAR data

of possible states within the model, and improves the accuracy of state estimations. Choosing more patterns of image blocks may further improve classification accuracy, however the computational complexity would be much larger as a result. Experimental results show 16 basic image block patterns results in relatively higher accuracy with lower computational complexity. Comparison for one of the classified images is shown in Figs. 4(c) and 4(d), the proposed 2D DHMM model has largely reduced the error rate of segmentation, both visually and quantitatively.

5. APPLICATION II: DHMM-BASED VIDEO CLASSIFICATION USING MULTIPLE INTERACTING MOTION TRAJECTORIES

In this section, we report experimental results of the proposed DHMM model applied to the task of multiple-object motion trajectory-based video classification. We test the classification performance of both proposed distributed 2D HMM-based classifier, causal 2D HMM-based classifier and traditional 1D HMM-based classifier on 2 datasets: (A) Synthetic multiple- trajectory dataset. (B) A subset of the Context Aware Vision using Image-based Active Recognition (CAVIAR)(http://homepages.inf.ed.ac.uk/rbf/CAVIAR/), which contains real-world video clips of multiple trajectories with interactions.

The results are reported in terms of 3 criteria: (1) The average Receiver Operating Characteristics (ROC) curve. (2) The Area Under Curve (AUC). (3) Classification Accuracy. The Classification Accuracy is defined as : $P_{Accuracy} = 1 - |F|/|S|$, where $|F|$ represents the cardinality of the false positives set, and $|S|$ represents the cardinality of the whole dataset.

We firstly test on the Synthetic dataset, whereas 1350 two-trajectory samples. We use 50% samples as training data, and the rest as testing data. The ROC curve is shown in Fig. 5. Test results show that our DHMM-based classifier achieves a 91.25% high accurate rate of classification on Synthetic dataset, shown in Table 1. We then test on the CAVIAR dataset, where we select data classes that have 2 people interacting with each other, and use 50% samples of ground truth trajectory as training data, keeping the rest as testing data. There are 9 classes of 180 two-people interacting trajectories. ROC curve is shown in Fig. 6. As shown in Table 1, the average classification accuracy of our DHMM-based classifier reaches 92.04%.

Table 1. Average Classification Accuracy Rates

Method–Dataset	SYNTHETIC(1350)	CAVIAR(180)
1D HMM	0.7654	0.8097
Causal 2D HMM	0.8319	0.8420
DHMM	**0.9125**	**0.9204**

6. CONCLUSION

In this paper, a novel multi-dimensional distributed hidden Markov model (DHMM) has been proposed. The proposed DHMM model provides an analytic solution to the non-causal multi-dimensional hidden Markov model by decomposing it into multiple distributed casual multi-dimensional hidden Markov models (HMMs). Simulation results in real-world image and video retrieval demonstrate the superior performance, higher accuracy rate and promising applicability of our DHMM model in comparison to previous models.

7. REFERENCES

[1] G. Soules L. E. Baum, T. Petrie and N. Weiss, “A maximization technique occuring in the statistical analysis of probabilitic functions of markov chains,” *Ann. Math. Stat.*, vol. 41(1).

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

[3] N. M. Laird A. P. Dempster and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *J. R. Stat. Soc.: Series B*, vol. 39(1), pp. 1–38, 1977.

[4] D. Schonfeld and N. Bouaynaya, “A new method for multidimensional optimization and its application in image and video processing,” *IEEE Sig. Proc. Letters*, vol. 13.