FACE RECOGNITION USING PSEUDO-2D ERGODIC HMM

S. A. Santosh Kumar D. R. Deepti B. Prabhakar

Central Research Laboratory, Bharat Electronics Limited, Bangalore-560 013, India.
as_santosh@yahoo.com, deepti_d_r@yahoo.com, prabhak_iitk@rediffmail.com

ABSTRACT

The work presented in this paper describes a novel Pseudo-2D Ergodic Hidden Markov Model (EHMM) based architecture for automatic face recognition. The primary HMM of this model being ergodic in nature, gives the flexibility to switch between the states, contrary to conventional Pseudo-2D HMM, which follows a top-to-bottom approach. The new approach helps in better modeling the different variations of a human face. We present a segmental K-means algorithm for training the Pseudo-2D EHMM, thereby jointly optimizing the observation densities and the state transitions corresponding to different variations of the face. The performance of the proposed method is presented with Discrete Cosine Transform (DCT) and the DCT-mod2 feature sets for the Olivetti Research Laboratory (ORL) database. The better modeling capability of the proposed architecture along with the robustness of DCT-mod2 feature set to illumination direction changes, proves to be an excellent combination for automatic face recognition.

1. INTRODUCTION

Face recognition from still images and video sequences is emerging as an active research area with numerous commercial and law enforcement applications. These applications require robust algorithms for human face recognition under different lighting conditions, facial expressions and orientations [1].

Various approaches for face recognition [2] include feature-based methods, template-based methods and the most recent ones being the model-based methods [1], [3], [4]. In [1], the potential of HMM to model human faces for identification purposes has been exploited. Faces can be intuitively divided into regions such as mouth, eyes, nose etc and these regions can be associated with the states of an HMM. The state observation densities of HMM can be represented using either of the two models namely, Gaussian Mixture Model (GMM) or Hidden Markov Model (HMM). In [1], GMM has been used to represent state observation densities of HMM. In [3], [4] HMM is used instead of GMM to model the state observation densities. The main HMM is referred to as primary HMM and the HMM used for modeling the observation densities as secondary HMM. Moreover in [3], [4], primary HMM is used to model two-dimensional data along one direction, and secondary HMM to model the data along the other direction. Both the HMMs follow a top-to-bottom approach.

In this paper, we present a Pseudo-2D HMM with state observation density modeled by a secondary HMM. The primary HMM is not restricted to follow a top-to-bottom approach, thereby giving flexibility to capture the different expressions and orientations of a face. This type of architecture has been recently used in language identification task [5]. Feature vectors for this work are obtained using 2D-DCT and a recent method termed as DCT-mod2 [6]. The compression and de-correlation property of 2D-DCT makes it suitable to use as feature vectors [3]. Moreover, the DCT-mod2 feature set derived from the basic 2D-DCT being robust to illumination direction changes, proves to be an added advantage for use as a feature set. We also present a segmental K-means algorithm for training the proposed Pseudo 2D-EHMM and report its performance for ORL database.

In the rest of this paper, the Pseudo-2D EHMM face recognition system is briefly described with results. Section 2 describes a EHMM based face recognition system and Section 3 explains the Pseudo-2D EHMM parameters. The training of the proposed Pseudo-2D EHMM by Segmental K-means algorithm is described in section 4. Section 5 deals with the experimental procedure and results. Conclusion of the work is given in section 6.

2. ERGODIC-HMM (EHMM) BASED FACE RECOGNITION SYSTEM

Fig. 1 shows a typical EHMM based face recognition system for N faces. An N – face recognition task is to classify an input face as belonging to one of 'N' faces, i.e., $F_1, F_2, \ldots, F_i, \ldots, F_N$. The EHMM system has N paths for a N face recognition task. A path 'i' (i = 1,…, N), has an EHMM ε_i of face F_i. For a given input face, the EHMM system yields N 'Viterbi likelihood' scores (V_i in Fig. 1), one for each F_i. The maximum – likelihood (ML) classifier identifies the face F_{i^*}, which has the highest likelihood score V_i, i.e., $i^* = \underset{i=1,\ldots,N}{arg\ max}\ V_i$.

3. EHMM PARAMETERS

The elements of Pseudo 2D-EHMM ε_i (Fig. 2) of face F_i, are as follows

E1. ***M*** **-** the number of states of the primary EHMM.

E2. **Observation density** $\boldsymbol{B}_i$: Observation densities are represented by secondary HMMs. Thus B_i is the set of M observation densities and is given by $B_i = \left\{ b_m^i(o) \right\}_{m=1}^{M} = \left\{ \lambda_1^i, \lambda_2^i, \ldots, \lambda_m^i, \ldots, \lambda_M^i \right\}$

where λ_m^i is the secondary HMM modeling the state 'm' and 'o' is the input feature vector sequence.

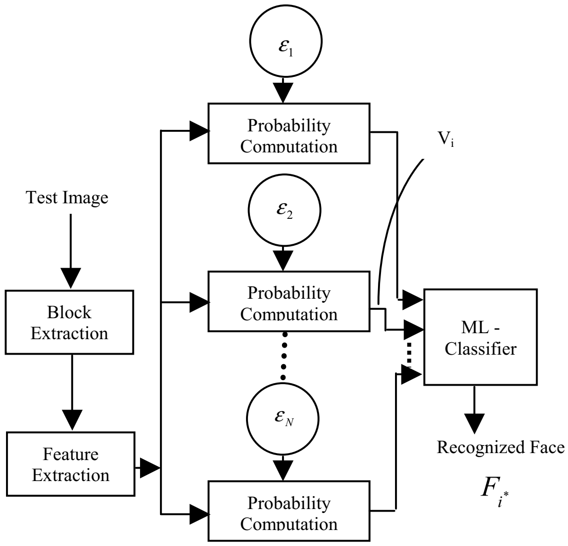

Fig. 1 Face Recognition by Ergodic HMM

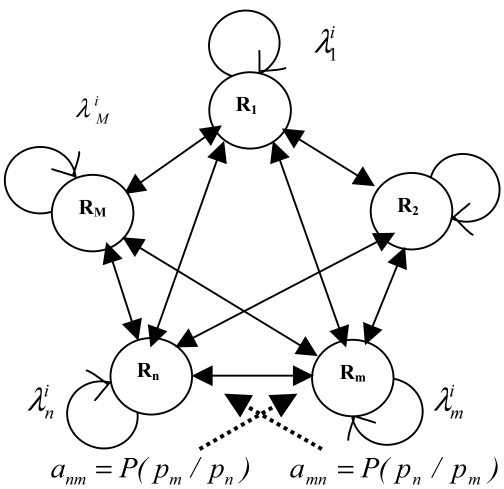

Fig. 2 Pseudo – 2D EHMM

E3. **Transition matrix A_i:** $A_i = \{a_{mn}\}$, m, n = 1,…, M specifies the transition probabilities a_{mn} between the states m and n covering the region R_m and R_n, which are modeled by secondary HMMs.

E4. **Initial state distribution π_i**: $\pi_i = \{\pi_{im}\}$, m = 1,..., M specifies π_{im}- the probability that state 'm' modeled by λ_m^i is the starting state.

An M state Ergodic HMM ε_i of face F_i using short hand notation is defined as triplet $\varepsilon_i = (A_i, B_i, \pi_i)$.

3.1 Viterbi likelihood

Given an input face feature vector $\mathbf{O} = (\mathbf{o_1},\mathbf{o_2},\ldots,\mathbf{o_t}\ldots,\mathbf{o_T})$, where T is the total number of feature vectors, the EHMM ε_i of face F_i can evaluate the Viterbi likelihood(score) V_i as,

$$
\begin{aligned}
V_i &= P^*(O,q \mid \varepsilon_i) = \max_q P(O,q \mid \varepsilon_i) \\
&= \max_q \{ P(O \mid q,\varepsilon_i) \,.\, P(q \mid \varepsilon_i)\} \\
&= \max_{q,B,K} \left\{ \pi_{iq1} p(s1 \mid \lambda_{q1}^i) \,.\, \prod_{k=2}^{K} \left[a_{q_{k-1}q_k} . p(s_k \mid \lambda_{qk}^i) \right] \right\} \qquad \ldots(1)
\end{aligned}
$$

where, B = $(b_o, b_1, \ldots, b_K)$, with b_o=0 and b_K = T , are the segment boundaries which segments $\mathbf{O}$ = $(\boldsymbol{o_1 o_2 o_3,\ldots,o_T})$ into K segments $(s_1, s_2, \ldots, s_k)$, where, segment $s_k = (o_{b_{k-1}+1}, \ldots, o_{b_k})$. q is any arbitrary state sequence of ε_i given by $q = (q_1, q_2, \ldots, q_{k-1}, q_k, \ldots, q_K)$ where state $q_k \in \{1,2,\ldots, M\}$. The corresponding observation density is $\lambda_{q_k}^i$ (drawn from $B_i = (\lambda_1^i, \lambda_2^i, \ldots, \lambda_M^i)$), which evaluates the probability of the 'observation segment' s_k, $b_{q_k}(s_k)$, as the Viterbi likelihood $p(s_k \mid \lambda_{q_k}^i)$.

Eqn. (1) maximizes V_i over the variables (q, B, K).

4. SEGMENTAL K-MEANS TRAINING OF EHMM

The parameters (A_i, B_i, π_i) of EHMM ε_i are learnt from the training faces, which contain different instances of the face F_i, i.e., $T_i = \{F_{ij}\}_{j=1}^{J}$, using a segmental K-means (SKM) algorithm. Initially each face image F_{ij} of width W and height H is divided into either overlapping or non-overlapping blocks of size P×Q. After extracting the blocks from each image in the training set, the feature vectors **O** obtained by DCT are used to train the model ε_i. Through the SKM, we jointly optimize the state observation densities and the state-transitions. Fig. 3 illustrates this procedure, which is as follows.

Step 1: Initialization

Set iteration count r = 1. Initialize

ε_i(r) = (A_i, B_i, π_i) with

1. A_i as equiprobable, i.e., $a_{mn} = 1/M$, $\forall$ m,n.
2. π_i as equiprobable, i.e., $\pi_{im} = 1/M$, m = 1,…, M.
3. Initialization of the state observation densities B_i:
Feature vector sequence belonging to each of the image in the training set is uniformly segmented into M regions taking into account the implicit notion that each state of HMM represents distinct regions of the face image like eyes, nose, mouth etc. The segments belonging to each of the M regions are modeled by HMM resulting in an inventory of M secondary HMMs, $\{b_m^i(o)\}_{m=1}^{M} = (\lambda_1, \lambda_2, \ldots, \lambda_M)$ (as in Fig. 3(a)).

Step 2: Viterbi decoding

Given a set of training sequences of face F_i, let $\mathbf{O} = (\mathbf{o_1 o_2},\ldots,\mathbf{o_T})$ be the feature vector sequences (of length T) of a typical face sequence F_{ij}. The Viterbi likelihood of **O** by ε_i(r) yields the Viterbi likelihood $P_{ij}^*(r)$ given by,

$$P_{ij}^*(r) = P^*(O, q \mid \varepsilon_i(r)) = \max_q \{P(O \mid q, \varepsilon_i(r)) P(q \mid \varepsilon_i(r))\} \qquad ...(2)$$

which is evaluated as in Eqn. (1).

Step 3: Parameter updation

Let $S_i(r) = \{S_{ij}(r)\}_{j=1}^J$ be the set of optimal state sequences obtained by Viterbi decoding of the training sequences $T_i = \{F_{ij}\}_{j=1}^J$ at iteration r of the SKM algorithm, i.e., $S_{ij}(r)$ is the optimal state sequence $q^* = (q_1^* q_2^*, ..., q_{K^*}^*)$ obtained by Viterbi decoding of feature vector sequence F_{ij} using ε_i (r) (the ergodic – HMM parameters at iteration (r)) as given by Eqn. (2). At iteration (r+1), the parameters (A_i, B_i, π_i) of ε_i (r+1) are updated using both $T_i = \{F_{ij}\}_{j=1}^J$ and $\{S_{ij}(r)\}_{j=1}^J$ as follows:

Update of A_i and π_i:

$$a_{mn} = n\left(q_{k-1}^* = m, q_k^* = n\right) / n\left(q_{k-1}^* = m\right) \quad m,n = 1,...,M \qquad ...(3)$$

$$\pi_{im} = n\left(q_1^* = m\right) / J \qquad m = 1,...,M \qquad ...(4)$$

where, the occurrence counts n (…) and n (.) are measured over all the J optimal state-sequences q^* of the J training sequences, i.e., over all the J sequences $\{S_{ij}(r)\}_{j=1}^J$.

Update of B_i:

Let $S_m = \{s_k : q_k^* = m\}$ be the set of all segments of $\{F_{ij}\}_{j=1}^J$ which have been assigned to state m (modeled by λ_m^i) by the optimal Viterbi decoding of Eqn. (2). Update the model λ_m^i using the segments in S_m, i.e., build a new HMM λ_m^i from these segments:

$$\lambda_m^i = HMM\left(S_m = \{s_k : q_k^* = m\}\right) \qquad m = 1,...,M \qquad ...(5)$$

Step 4: Convergence

$P_i^*(r) = \frac{1}{J}\sum_{j=1}^J P_{ij}^*(r) / T_j$ is the average Viterbi likelihood over all the training sequences of face F_i, after Eqn. (2) of iteration r. T_j is the total number of vectors in feature vector set for face F_{ij} and $P_{ij}^*(r)$ is the Viterbi likelihood as per Eqn. (2) using ε_i(r) at iteration r.

Terminate SKM iteration if $\left| P_i^*(r) - P_i^*(r-1) \right| < \delta$; otherwise continue with 'Step 2: Viterbi decoding' with r = r+1; δ is a suitable threshold to ensure a good convergence.

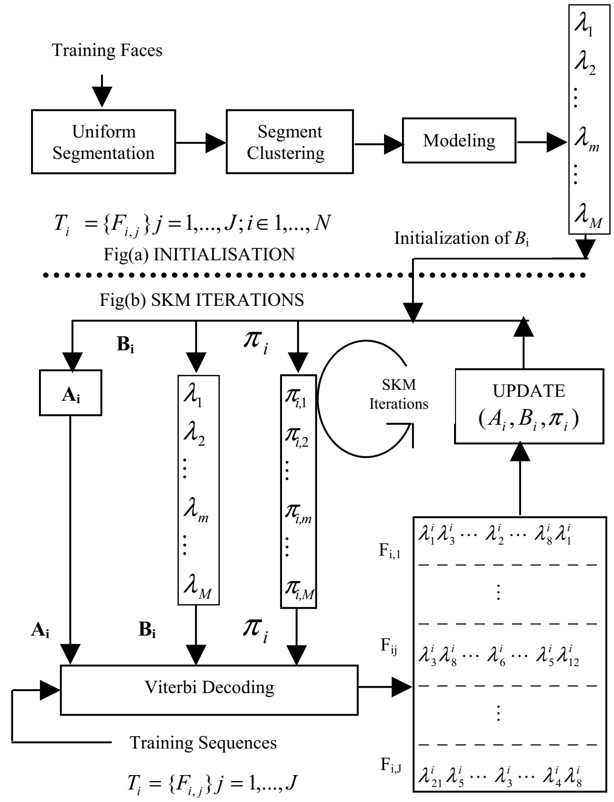

Fig. 3 EHMM Training By SKM
Fig. 3(a) Generation of Face Dependent Inventory
Fig. 3(b) SKM Iterations for EHMM Parameter Estimation of Face F_i

5. EXPERIMENTS AND RESULTS

We present here experimental results of face recognition using the proposed Pseudo-2D EHMM system on ORL database.

5.1 Database

The proposed Pseudo 2D-EHMM performance is evaluated on a set of 40 faces of the ORL database, with 10 different images per face. The system is trained on 6 images and tested on the 4 remaining images available per face.

5.2 Feature extraction

Each face image $F_{i,j}$ is divided into overlapping or non-overlapping blocks of size P×Q, P and Q are chosen to be equal to 8 for this work. The proposed face recognition algorithm is tried out on three different feature sets, viz., 2D-DCT baseline without overlap, 2D-DCT baseline with overlap and DCT-mod2, which are described below.

i. '2D-DCT baseline without overlap' feature vector set $\mathbf{O} = \{\mathbf{o_1, o_2, ..., o_t, ..., o_T}\}$ is formed by extracting 2D-DCT coefficients for the blocks 1,2…t,…T. Each feature vector $\mathbf{o_t}$ in the feature set $\mathbf{O}$ is formed after zigzag scanning of the extracted DCT coefficients for the block 't' and is given by $o_t = [c_0, c_1, \cdots, c_d]'$.

ii. The same procedure is repeated with 50% overlap between the blocks to form the '2D-DCT baseline with overlap' feature set.

iii. 'DCT-mod2 with overlap' feature set is formed by considering the horizontal and vertical neighbours of each block. DCT-mod2 feature vector for a block 't' located at (x, y), is given as below,

$$o_{t,(x,y)} = \left[\Delta^h c_0 \Delta^v c_0 \Delta^h c_1 \Delta^v c_1 \Delta^h c_2 \Delta^v c_2 c_3 ... c_{d-4}\right]'.$$

It is formed from the '2D-DCT baseline with overlap, by replacing the first three coefficient values by their horizontal and vertical delta values, which are given by,

$$\Delta^h c_n^{(x,y)} = \sum_{k=-1}^{1} k c_n^{(x,y+k)} / 2$$
$$\Delta^v c_n^{(x,y)} = \sum_{k=-1}^{1} k c_n^{(x+k,y)} / 2 \quad ; \quad n = 0, 1, 2$$

The dimensionality '*d*' of feature vector for all the feature sets is chosen to be 16.

5.3 Parameters of EHMM

In EHMM based face recognition system, the main parameters are the number of states of the primary HMM '*M*', the number of states of the secondary HMM and the number of Gaussian mixtures/state of secondary HMM. In our experiment, for each face, EHMM systems were designed for M = 1,3,5,7,11, and each state of primary HMM modeled by 3 state left to right secondary HMM with 3 Gaussian mixtures/state.

5.4 Results

	NUMBER OF STATES (*M*)				
Feature Set	1	3	5	7	11
2D-DCT Without Overlap	93.3	100	100	100	100
2D-DCT With Overlap	91.6	100	100	100	100
DCT –mod2 With Overlap	92.49	100	100	100	100

Table 1. Training Data Accuracy in %

	NUMBER OF STATES (*M*)				
Feature Set	1	3	5	7	11
2D-DCT Without Overlap	80.0	93.13	93.13	**95.0**	93.75
2D-DCT With Overlap	75.63	92.5	97.5	**98.13**	97.5
DCT – mod2 With Overlap	77.5	95.63	97.5	**100**	98.13

Table 2. Test Data Accuracy in %

Tables 1 and 2 show the training and test data performance respectively of the proposed face recognition system on ORL database with the different feature sets chosen for a number of states '*M'* varying from 1 to 11.

Table 1 shows the potential of the proposed architecture in achieving very high face recognition accuracy. It has shown an accuracy of 100% with all the feature sets except for M=1.

Table 2 shows the test data performance, indicating an increase in performance with increase in M up to M=7. Recognition accuracy has shown its maximum for M=7 with all feature sets, giving a 100% accuracy with DCT-mod2. The reduction in the test data performance for M greater than 7 is possibly due to shortage of training data to get generalized to the test data or may be an over fitting problem, i.e., the model becomes too training data specific that it fails to recognize the test data.

The previous literature on face recognition using embedded or Pseudo-2D HMM [3] has shown a maximum accuracy of 98% on 'DCT baseline with overlap'. Our architecture using 'DCT baseline with overlap' has achieved similar performance by giving an accuracy of 98.13%.

6. CONCLUSIONS

A novel scheme for face recognition namely Pseudo-2D EHMM, based on modeling of the state observation density by an HMM, is proposed in this work. The training algorithm for the proposed system is presented and its performance is evaluated using ORL database consisting of 40-faces. Since, the proposed pseudo-2D EHMM architecture does not follow the conventional top-to-bottom approach, it is found to be efficient in capturing the varying expressions and orientations of faces. Further more DCT-mod2 features set is proved to be robust to illumination direction changes. The proposed architecture using a combination of the above has resulted in a highly accurate and efficient face recognition scheme.

ACKNOWLEDGEMENTS

The first author is indebted to Prof. T. V. Sreenivas and Dr. V. Ramasubramanian, Dept of ECE, Indian Institute of Science (IISc), Bangalore, India, for exposing him to the world of HMM. The authors are grateful to Dr. R. Rajagopal, SRS (Radar Signal Processing group (RSP), Central Research Laboratory (CRL)), Bangalore, India, for the constant encouragement. The authors are also thankful to C. R. Patil (Member (Research staff)), CRL, for reviewing and giving useful comments.

REFERENCES

[1] A.V. Nefian and M. H. Hayes III, "A hidden Markov model for face recognition," In *Proc. ICASSP*, vol.5, pp. 2721-2724, 1998.

[2] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," In *Proc. of IEEE*, vol. 83, No. 5, pp. 705-740, 1995.

[3] A. V. Nefian and M. H. Hayes III, "An embedded HMM – based approach for face detection and recognition," In *Proc. ICASSP*, vol.6, pp. 3553-3556, 1999.

[4] A. V. Nefian and M. H. Hayes III, "Maximum likelihood training of the embedded HMM for face detection and recognition," In *Proc. ICIP,* Vol. 1, pp 33-36, 2000.

[5] S. A. Santosh Kumar and V. Ramasubramanian, "Automatic language identification using Ergodic – HMM," In *Proc. ICASSP*, vol.1, March 14-23, pp. 609-612, 2005.

[6] C. Sanderson and K. K. Paliwal, "Fast features for face authentication under illumination direction changes," Pattern Recognition letters, Vol. 24, No. 14, pp. 2409-2419, 2003.