HIDDEN MARKOV MODELS FOR FACE RECOGNITION

Ara V. Nefian and Monson H. Hayes III

Center for Signal and Image Processing
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA 30332
{ara, mmh3}@eedsp.gatech.edu

ABSTRACT

The work presented in this paper focuses on the use of Hidden Markov Models for face recognition. A new method based on the extraction of 2D-DCT feature vectors is described, and the recognition results are compared with other face recognition approaches. The method introduced in this paper reduces significantly the computational complexity of previous HMM-based face recognition system, while preserving the same recognition rate.

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. An excellent survey on face recognition is given in [1]. Previous attempts to develop a face recognition system that has a high recognition rate include the correlation method [2], the eigenface method [3] and the linear discriminant method [4]. However, the recognition rate in each of these methods decreases rapidly when the face orientation or the face image size change. In order to avoid these problems, for each of these methods a view based approach was developed [5]. In the first stage of these methods, the orientation, or facial expression is determined, and then the recognition is performed using the database corresponding to images that have the given orientation or facial expression.

Given the success of HMMs in speech recognition and character recognition, and the work of Samaria [6], we have developed a face recognition system using HMM. In this paper we describe this face recognition system and compare the recognition results to the eigenface method and the earlier HMM based approach.

2. HIDDEN MARKOV MODELS

Hidden Markov Models (HMM) are a set of statistical models used to characterize the statistical properties of a signal [7]. HMM consist of two interrelated processes: (1) an underlying, unobservable Markov chain with a finite number of states, a state transition probability matrix and an initial state probability distribution and (2) a set of probability density functions associated with each state. The elements of a HMM are:

- N, the number of states in the model. If S is the set of states, then $S = \{S_1, S_2, ..., S_N\}$. The state of the model at time t is given by $q_t \in S$, $1 \le t \le T$, where T is the length of the observation sequence (number of frames).
- M, the number of different observation symbols. If V is the set of all possible observation symbols (also called the *codebook* of the model), then $V = \{v_1, v_2, ..., v_M\}$.
- $\mathbf{A}$, the state transition probability matrix, i.e. $\mathbf{A} = \{a_{ij}\}$ where

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i] \; 1 \le i, j, \le N, \tag{1}$$

  with the constraint,

$$0 \le a_{i,j} \le 1,$$

  and,

$$\sum_{j=1}^{N} a_{ij} = 1, \; 1 \le i \le N$$

- $\mathbf{B}$, the observation symbol probability matrix, i.e. $\mathbf{B} = \{b_j(k)\}$, where

$$b_j(k) = P[\mathbf{O_t} = v_k | q_t = S_j], \tag{2}$$

$$1 \le j \le N, \; 1 \le k \le M$$

  and $\mathbf{O_t}$ is the observation symbol at time t.
- $\mathbf{\Pi}$, the initial state distribution, i.e. $\mathbf{\Pi}=\{\pi_i\}$ where:

$$\pi_i = P[q_1 = S_i], \; 1 \le i \le N \tag{3}$$

Using a shorthand notation, a HMM is defined as the triplet

$$\lambda = (\mathbf{A}, \mathbf{B}, \mathbf{\Pi}). \tag{4}$$

The above characterization corresponds to a discrete HMM, where the observations are discrete symbols chosen from a finite alphabet $V = \{v_1, v_2, ..., v_M\}$. In a continuous density HMM, the states are characterized by continuous observation density functions. The most general representation of the model probability density function (pdf) is a finite mixture of the form:

$$b_i(\boldsymbol{O}) = \sum_{k=1}^{M} c_{ik} N(\boldsymbol{O}, \mu_{ik}, U_{ik}),\ 1 \leq i \leq N \tag{5}$$

where c_{ik} is the mixture coefficient for the kth mixture in state i. Without loss of generality $N(\mathbf{O}, \mu_{ik}, U_{ik})$ is assumed to be a Gaussian pdf with mean vector μ_{ik} and covariance matrix U_{ik}.

3. FACE IMAGE HMM

Hidden Markov Models have been successfully used for speech recognition where data is essentially one dimensional. Extension to a fully connected two dimensional HMM has been shown to be computationally very complex [8]. In [9], Kuo and Agazzi have used a pseudo two dimensional HMM for character recognition that was shown to perform reasonably fast for binary images. In this paper we investigate the recognition performance of a one dimensional HMM for gray scale face images.

For frontal face images, the significant facial regions (hair, forehead, eyes, nose, mouth) come in a natural order from top to bottom, even if the images are taken under small rotations in the image plane and/or rotations in the plane perpendicular to the image plane. Each of these facial regions is assigned to a state in a left to right 1D continuous HMM. The state structure of the face model and the non-zero transition probabilities a_{ij} are shown in Figure 1.

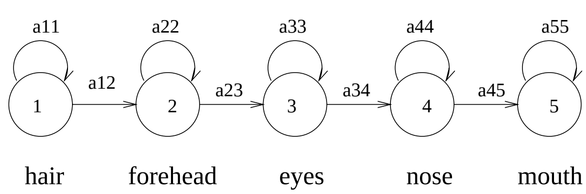


Figure 1: Left to right HMM for face recognition

4. FEATURE EXTRACTION

Each face image of width W and height H is divided into overlapping blocks of height L and width W. The amount of overlap between consecutive blocks is P (Figure 2).

The number of blocks extracted from each face image equals the number of observation vectors T and is given by:

$$T = \frac{H - L}{L - P} + 1, \tag{6}$$

The choice of parameters P and L can significantly affect the system recognition rate. A high amount of overlap P significantly increases the recognition rate because it allows the features to be captured in a manner that is independent of the vertical position. The choice of parameter L is more delicate. Using a small L can bring insufficient discriminant information to the observation vector, while large L increases the probability of cutting across the features.

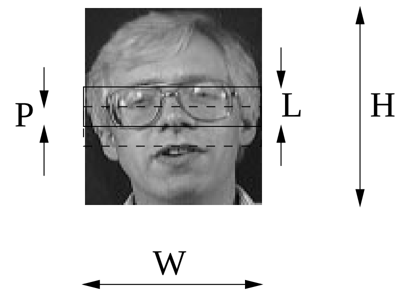


Figure 2: Face image parameterization and blocks extraction

However, the system recognition rate is not very sensitive to variations in L, as long as P is large ($P \leq L - 1$). In [10] the effect of parameters P and L, together with the effect of the number of states used in the HMM has been extensively discussed.

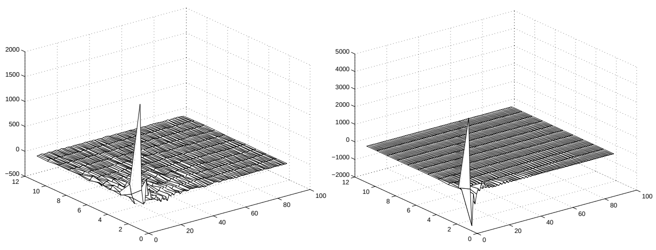
In [6] the observation vectors consist of all the pixel values from each of the blocks, and therefore the dimension of the observation vector is $L \times W$ ($L = 10$ and $W = 92$). The use of the pixel values as observation vectors has two important disadvantages: First, pixel values do not represent robust features, being very sensitive to image noise as well as image rotation, shift or changes in illumination and second, the large dimension of the observation vector leads to high computational complexity of the training and recognition systems, and therefore increases the processing time required for recognition. This can be a major problem for face recognition over large databases or when the recognition system is used for real time applications. In this paper, the observation vectors consist of a set of 2D-DCT coefficients that are extracted from each block. The DCT compression properties for natural images make the use of this transform a suitable feature extraction technique for the face recognition system. The typical DCT coefficients for significant facial regions are shown in Figure 3. The features used in this approach, contain the 2D-DCT coefficients inside a rectangular window of size 13×3 over the lowest frequencies in the DCT domain. The window size has been chosen to cover the most significant coefficients, i.e. the coefficients that contain most of the signal energy.

5. TRAINING THE FACE MODELS

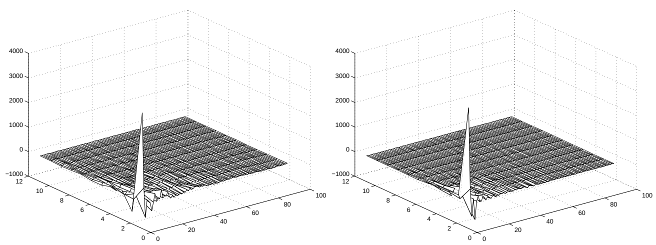
Each individual in the database is represented by a HMM face model. A set of five images representing different instances of the same face are used to train each HMM. Following the block extraction, a set of 39 2D-DCT coefficients obtained from each block are used to form the observation vectors (Figure 4). The observation vectors are effectively used in the training of each HMM.

First, the HMM $\lambda = (\mathbf{A}, \mathbf{B}, \mathbf{\Pi})$ is initialized. The training data is uniformly segmented from top to bottom in $N = 5$ states and the observation vectors associated with each state are used to obtain initial estimates of the observation probability matrix $\mathbf{B}$. The initial values for $\mathbf{A}$ and $\mathbf{\Pi}$ are set given the left to right structure of the face model.

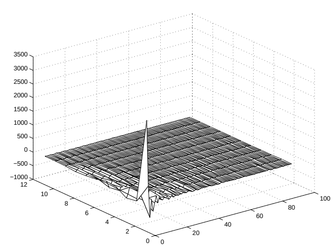
In the next steps the model parameters are re-estimated



(a)



(b)



(c)

Figure 3: Typical DCT coefficients for: a - hair (left) and forehead (right), b - eyes (left) and nose (right), c - mouth.

using the E-M procedure [11] to maximize $P(\mathbf{O}|\lambda)$. The iterations stop, after model convergence is achieved, *i.e* the difference between model probability at consecutive iterations (k and $k+1$) is smaller than a threshold C,

$$|P(\boldsymbol{O}|\lambda^{(k+1)}) - P(\boldsymbol{O}\ |\lambda^{(k)})| < C \qquad (7)$$

6. RECOGNITION AND RESULTS

In the recognition phase, a set of 200 test images, not used in the training, are considered to determine the recognition performances of the system. After extracting the observation vectors as in the training phase, the probability of the observation vector given each HMM face model is computed. A face image t is recognized as face k if:

$$P(\boldsymbol{O}^{(t)}|\lambda_k) = max_n P(\boldsymbol{O}^{(t)}|\lambda_n) \qquad (8)$$

The face recognition system has been tested on the Olivetti Research Ltd. database (400 images of 40 individuals, 10 face images per individual at the resolution

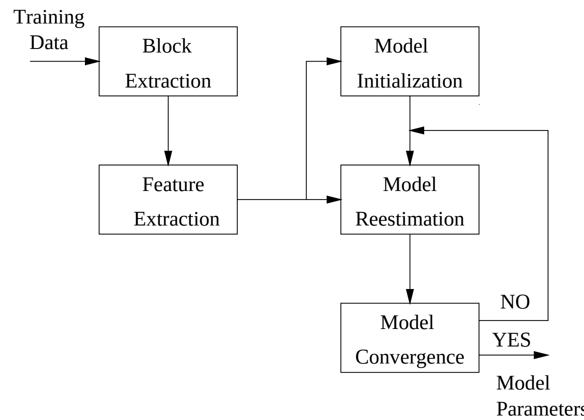


Figure 4: HMM training scheme

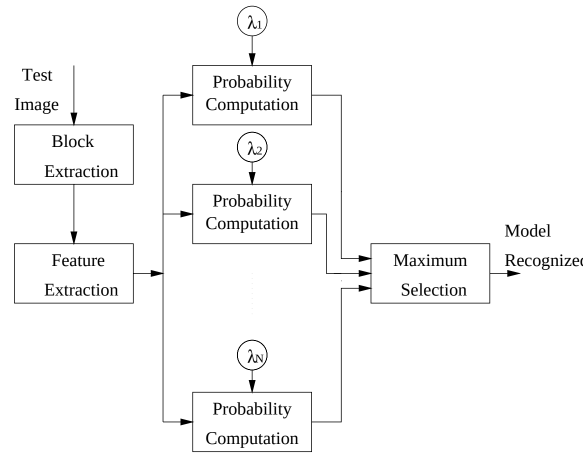


Figure 5: HMM recognition scheme

of 92 × 112 pixels). The database contains face images showing different facial expressions, hair styles, eye wear (glasses/no glasses), and head orientations. The system achieved a recognition rate of 84% with $L = 10$ and $P = 9$. On the same database the recognition rate of the eigenface method is 73% and the recognition rate of the HMM based approach presented in [6] is 84% over a fraction of the same database. The approach presented in this paper performs at a recognition rate better than the eigenface method and decreases significantly the recognition time of the HMM based method in [6] while preserving the same recognition rate. The processing time required to compute the likelihood of one test image given a face model is decreased from 25 seconds reported in [6] (C language processing time) to 2.5 seconds in the present work (Matlab processing time).

Figure 6 presents some of the recognition results. The crossed images represent incorrect classifications, while the rest of images are examples of correct classification. The horizontal lines show the state segmentation that was obtained using the Viterbi algorithm. It can be noticed that the state segmentation of the HMM face model separates the significant facial regions such as hair/forehead, eyes, nose and mouth.

7. CONCLUSIONS

This paper describes a HMM based approach for face recognition that uses 2D-DCT coefficients as feature vectors. The method is compared to the earlier HMM-based face recognition system in [6], where the pixel values of each block form the observation vectors, and with the classical eigenface method. Both HMM based approaches show the same recognition performance and better recognition rates than the eigenface method. Due to the compression properties of the DCT, the size of the observation vector in the current approach is reduced from $L \times W$ ($L = 10$ and $W = 92$) to 39, while preserving the same recognition rate (The recognition performance in [6] was based on a fraction of the database, while the experiments presented here were conducted over all images of the same database). The use of a lower dimensional feature vector (over 23 times smaller than the size of the observation vector in the previous method) leads to a significant reduction of the computational complexity of the method and consequently to a significant decrease of the face recognition time. The HMM modeling of human faces appears to be an encouraging method for face recognition under a wider range of image orientations and facial expressions. Future work will be directed on the study of pseudo 2D HMM for face image recognition, the inclusion of state duration in face modeling, as well as other feature extraction techniques.

8. REFERENCES

[1] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," Proceedings of IEEE, vol. 83, May 1995.

[2] D. Beymer, "Face recognition under varying pose," in Proceedings of 23rd Image Understanding Workshop, vol. 2, pp. 837–842, 1994.

[3] M. Turk and A. Pentland, "Face recognition using eigenfaces," in Proceedings of International Conference on Pattern Recognition, pp. 586 – 591, 1991.

[4] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs Fisherfaces: Recognition using class specific linear projection," in Proceedings of Fourth Europeean Conference on Computer Vision, ECCV'96, pp. 45–56, April 1996.

[5] A. Pentland, B. Moghadam, T. Starner, and M. Turk, "View based and modular eigenspaces for face recognition," in Proceedings on IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 84–91, 1994.

[6] F. Samaria and S. Young, "HMM based arhitecture for face identification," Image and Computer Vision, vol. 12, pp. 537–583, October 1994.

[7] L. Rabiner and B. Huang, Fundamentals of Speech Recognition. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[8] E.Levin and R. Pieraccini, "Dynamic planar warping for optical character recognition," in ICASSP, pp. 149–152, 1992.

[9] S. Kuo and O. Agazzi, "Keyword spotting in poorly printed documents using pseudo 2-d Hidden Markov Models," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994.

[10] F. Samaria and A. Harter, "Parametrisation of stochastic model for human face identification," in Proceedings of the Second IEEE Workshop on Application of Computer Vision, 1994.

[11] C. Wu, "On the convergence properties of the EM algorithm," Annals of Statistics, vol. 11, no. 1, pp. 95–103, 1983.

Figure 6: Recognition Results