FACE DETECTION AND RECOGNITION USING HIDDEN MARKOV MODELS

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 describes a Hidden Markov Model (HMM)-based framework for face recognition and face detection. The observation vectors used to characterize the states of the HMM are obtained using the coefficients of the Karhunen-Loeve Transform (KLT). The face recognition method presented in this paper reduces significantly the computational complexity of previous HMM-based face recognition systems, while slightly improving the recognition rate. Consistent with the HMM model of the face, this paper introduces a novel HMM-based face detection approach using the same feature extraction techniques used for face recognition.

1. INTRODUCTION

Face recognition and face detection, from still and video images is emerging as an active research area with numerous commercial and law enforcement applications. Face detection and recognition systems can be used to allow access to an ATM machine or a computer, to control the entry of people into restricted areas, and to recognize people in specific areas (banks, stores), or in a specific database (police database). A robust face identification system must operate under a variety of conditions, such as varying illuminations and backgrounds, it must be able to handle non-frontal facial images of both males and females of different ages and races, and be robust in the presence of two or more faces within an image.

A survey on face recognition is given in [1]. Previous attempts to develop a face recognition system include geometric feature-based methods, template-based methods ([2], [3], [4]), and more recently model-based methods ([5], [6]). In [5], Samaria introduced an HMM-based face recognition system. For a frontal face, the significant facial regions such as hair, forehead, eyes, nose and mouth occur in a natural order from top to bottom. Therefore, each of these facial regions is assigned to a state, in a left-to-right one-dimensional continuous HMM.

This paper presents a face recognition and detection system using an HMM approach. An efficient method for extracting the observation vectors using the KLT coefficients is presented. Consistent with the HMM framework for face recognition, a novel face detection approach is introduced. Compared to the classical template-based methods, the HMM-based approach offers a more flexible framework for detection and recognition, and can be used more efficiently in scale invariant systems.

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: (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 $q_t \in S$, $1 \leq t \leq T$, where T is the length of the observation sequence (number of frames).

- $\mathbf{\Pi}$, the initial state distribution, i.e. $\mathbf{\Pi}=\{\pi_i\}$ where:

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

- $\mathbf{A}$, the state transition probability matrix, i.e. $\mathbf{A} = \{a_{ij}\}$ where

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i]\ 1 \leq i, j, \leq N, \tag{2}$$

with the constraint,

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

and,

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

- $\mathbf{B}$, the state probability matrix, i.e. $\mathbf{B} = \{b_j(\mathbf{O}_t)\}$. 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:

$$b_i(\mathbf{O}_t) = \sum_{k=1}^{M} c_{ik} N(\mathbf{O}_t, \mu_{ik}, U_{ik}), \ 1 \leq i \leq N \quad (3)$$

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

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

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

3. FACE IMAGE HMM

Hidden Markov Models have been successfully used for speech recognition and more recently in action recognition where data is essentially one dimensional over time. In this paper we investigate the recognition and detection 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 undergo 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.

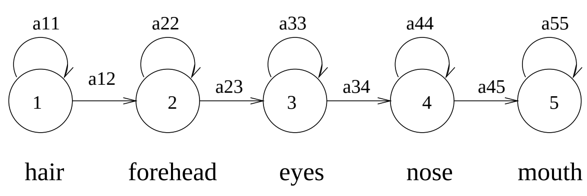


Figure 1: Left to right HMM for face recognition

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).

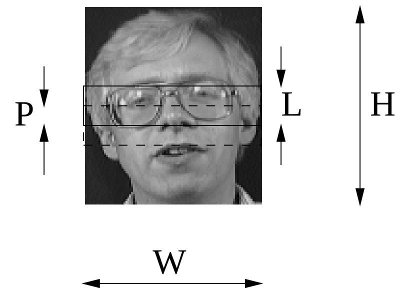


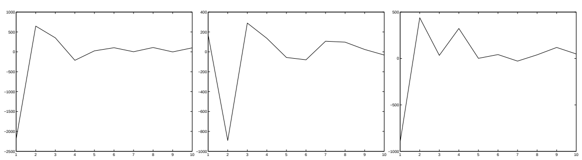
Figure 2: Face image parameterization and blocks extraction

The variations of the recognition performance with parameters P and L is extensively disscused in [5]. However, the performance of the system is less sensitive to variations in L, as long as P remains large ($P \leq L - 1$).

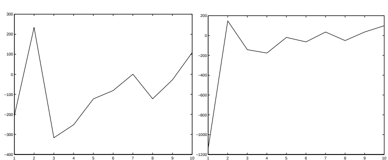
In [5] the observation vectors consist of all the pixel values from each of the blocks, and therefore the dimension of the observation vector is $L \times W$. The use of the pixel values as observation vectors has two important disadvantages: First, pixel values do not represent robust features, being very sensitive to image noise as well as image rotation, shift or changes in illumination. Second, the large dimension of the observation vector leads to high computational complexity of the training and detection/recognition systems. This can be critical for a face detection or recognition system that operates on a large database or when the detection/recognition system is used for real time applications.

In [6], the 2D-DCT coefficients extracted from each block were used to obtain an efficient set of observation vectors. The coefficients inside a rectangular window over the lowest frequencies in the DCT domain, which concentrate most of the block energy, were used as observation vectors. The choice of the 2D-DCT coefficients reduced dramatically the size of the observation vectors and therefore decreased the complexity of the system.

In this paper, the observation vectors consist of the Karhunen Loeve Transform (KLT) coefficients. The KLT compression properties as well as its decorrelation properties make it an attractive technique for the extraction of the observation vectors. From the images in the training set, the blocks are extracted and arranged column-wise to form vectors. The eigenvectors corresponding to the largest eigenvalues of the covariance matrix of these vectors form the KLT basis. Let μ be the mean of the vectors used to compute the covariance matrix. To obtain the observation vectors, the mean vector μ is subtracted from each of the vectors corresponding to a block in the image. The resulting vector is then projected onto the eigenvectors of the covariance matrix and the resulting coefficients form the observation vectors. In Figure 3 the typical observation vectors for blocks corresponding to some significant facial regions are illustrated.



(a)



(b)

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

5. TRAINING THE FACE MODELS

For face detection, a set of face images is used in the training of one HMM. The images in the training set represent frontal faces of different people taken under different illumination conditions.

For face recognition, each individual in the database is represented by an HMM face model. A set of images representing different instances of the same face are used to train each HMM.

After extracting the blocks from each image in the training set, the observation vectors (KLT coefficients) are obtained and used to train each of the HMMs. First, the HMM $\lambda = (\mathbf{A}, \mathbf{B}, \mathbf{\Pi})$ is initialized as follows. The training data is uniformly segmented from top to bottom in $N = 6$ 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. (Figure 4). At the next iteration, the uniform segmentation is replaced by the Viterbi segmentation. The iteration stop, and the HMM is initialized, when the Viterbi segmentation likelihood at consecutive iterations is smaller than a threshold. The final parameters of the HMM are obtained using the Baum-Welch recursive procedure.

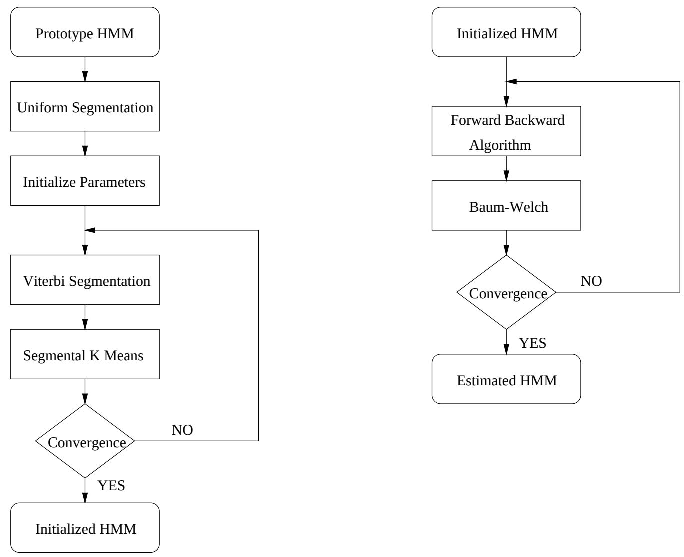


Figure 4: Training Scheme

6. FACE DETECTION

Given a test image that contains one or more faces in a cluttered background, face detection begins by looking within each rectangular window in the test image, extracting the observation vectors, and computing the probability of data inside each window given the face model. In order to cover all possible face locations in the image, adjacent windows have a large amount of overlap in both the horizontal and vertical directions. For simplicity, the probability computation is carried out via the Viterbi algorithm. The windows that have a face model likelihood higher than a threshold are selected as possible face locations. In order to remove the false alarms in windows close to the detected face position, only the windows with the maximum face likelihood over a vicinity around the current window location are selected as candidate faces. This approach allows a fast implementation due to the breaking of the face template into short observation vectors. The computation of a face likelihood of a window requires only the computation of a few new observation vectors, corresponding to the amount of shifting between consecutive windows, followed by the Viterbi segmentation, which is relatively inexpensive procedure (Figure 5).

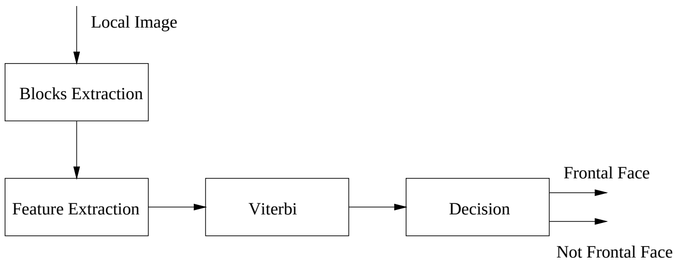


Figure 5: Face Detection

Figure 6 shows some of the face detection results. The horizontal lines show the Viterbi state segmentation. The face detection system has been tested on the MIT database (48 images of 16 people with background, the face resolution is about 60×90 pixels) taken under different illumination conditions. The manually segmented faces from 9 images were used in the training set. The other 39 images were used for testing. Our experiments show 90% correct detection over this database.

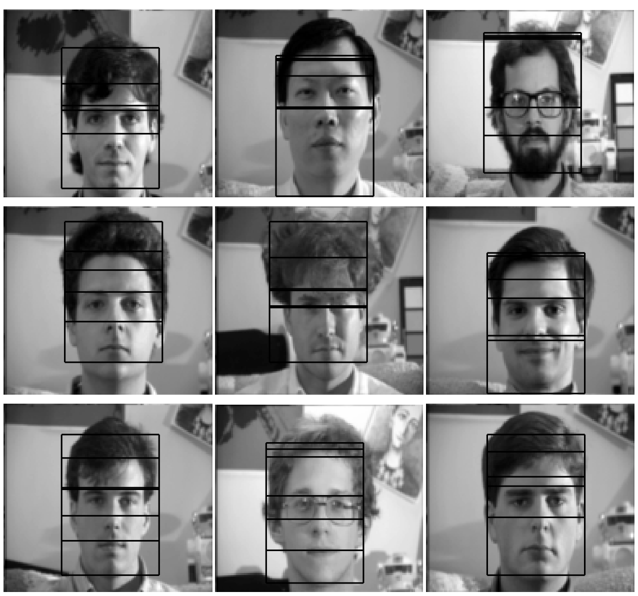


Figure 6: Face Detection Results

7. FACE RECOGNITION

The HMM face modeling capabilities for face recognition were tested using a set of face images (with no background) not used in the training phase.

After extracting the observation vectors as in the training phase, the probability of the observation sequence given each HMM face model is computed via a simple Viterbi recognizer. The model with the highest likelihood is selected and this model reveals the identity of the unknown face (Figure 7).

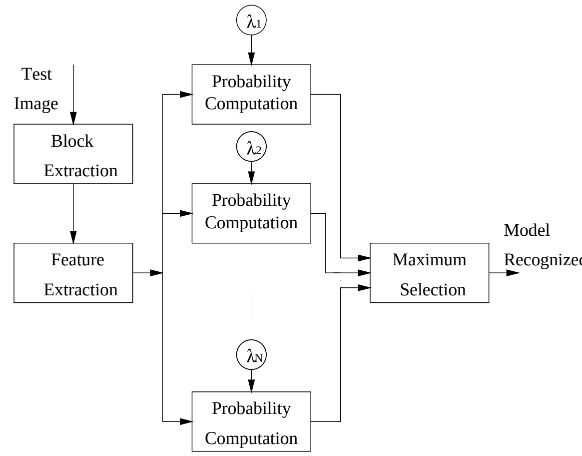


Figure 7: HMM recognition scheme

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 of 92×112 pixels). Half of the images were used in training, and the other half were used for testing. The database contains face images showing different facial expressions, hair styles, and eye wear (glasses/no glasses). On the same database the recognition performance of the method in [5] and [6] was 84%. The approach in [5] is computationaly very complex because of the large size of observation vectors. The accuracy of the system presented in this paper is slightly increased to 86 % while the recognition time is 250ms / face compared to 700ms / face in [6].

Figure 8 presents some of the recognition results. The crossed images represent incorrect classifications, while the rest of images are examples of correct classification.

8. CONCLUSIONS

This paper describes an HMM-based approach for face recognition and detection that uses an efficient set of observation vectors based on the extraction of the KLT coefficients.

A novel HMM-based face detection approach was introduced in this paper. The accuracy of this method with respect to variations in lighting conditions and its complexity efficiency, suggest that this method may be a promising approach for face detection.

The HMM modeling of human faces appears to be a promising method for face recognition and face de-

Figure 8: Face Recognition Results

tection under a wider range of image orientations and facial expressions.

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

[6] A. V. Nefian and M. H. Hayes, "Hidden markov models for face recognition," in *ICASSP98*, pp. 2721–2724, 98.

[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.