AN EMBEDDED HMM - BASED APPROACH FOR FACE DETECTION AND RECOGNITION

Ara V. Nefian and Monson H. Hayes III

Georgia Institute of Technology
School of Electrical and Computer Engineering
Center for Signal and Image Processing
Atlanta, GA, 30332
{ara, mhh3}@eedsp.gatech.edu

ABSTRACT

In this paper we describe an embedded Hidden Markov Model (HMM)-based approach for face detection and recognition that uses an efficient set of observation vectors obtained from the 2D-DCT coefficients. The embedded HMM can model the two dimensional data better than the one-dimensional HMM and is computationally less complex than the two-dimensional HMM. This model is appropriate for face images since it exploits an important facial characteristic: frontal faces preserve the same structure of "super states" from top to bottom, and also the same left-to-right structure of "states" inside each of these "super states".

1. INTRODUCTION

A face identification system can be used to detect the location of faces from different scenes and recognize them as one of the faces stored in a database. The system must operate under a variety of conditions, such as varying illuminations and backgrounds, and it must be able to handle non-frontal facial images of males and females of different ages and races.

Previous approaches to face recognition [1] include geometric feature-based methods, template-based methods, and more recently Hidden Markov Model (HMM) - based methods [2]. This last approach exploits the fact that the most significant facial features of a frontal face image occur 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. Therefore, the image of a face may be modeled using a one-dimensional HMM by assigning each of these regions to a state [2] [3] [4]. The one-dimensional HMM was extended by Samaria to a pseudo two-dimensional HMM by adding a *marker block* at the end of each line in the image, and introducing an additional *end-of-line state* at the end of each horizontal HMM [5] to preserve the two dimensional structure of the data.

In this paper, we describe a new approach to face recognition and detection using an embedded HMM as introduced by Kuo and Agazzi for character recognition [6]. The observation vectors used by our embedded HMM are obtained from the two-dimensional Discrete Cosine Transform (2D-DCT) coefficients.

2. AN EMBEDDED HMM

A one-dimensional HMM [7] may be generalized, to give it the appearance of a two-dimensional structure, by allowing each state in a one-dimensional *overall* HMM to be an HMM. In this way, the HMM consists of a set of *super states*, along with a set of *embedded states*. The super states may then be used to model two-dimensional data along one direction, with the embedded HMM modeling the data along the other direction. The elements of an embedded HMM are:

- The number of super states, N_0, and the set of super states, $\mathbf{S_0} = \{S_{0,i}\}\ 1 \leq i \leq N_0$.
- The initial super state distribution, $\mathbf{\Pi_0} = \{\pi_{0,i}\}$, where $\pi_{0,i}$ are the probabilities of being in super state i at time zero.
- The super state transition probability matrix,

  $$\mathbf{A_0} = \{a_{0,ij}\}$$

  where $a_{0,ij}$ is the probability of transitioning from super state i to super state i.
- The parameters of the embedded HMMs $\mathbf{\Lambda}$, which include
  - The number of embedded states in the k th super state, $N_1^{(k)}$, and the set of embedded states, $\mathbf{S}_1^{(k)} = \{S_{1,i}^{(k)}\}$.
  - The initial state distribution, $\mathbf{\Pi}_1^{(k)} = \{\pi_{1,i}^{(k)}\}$, where $\pi_{1,i}^{(k)}$ are the probabilities of being in state i of super state k at time zero.
  - The state transition probability matrix,

    $$\mathbf{A}_1^{(k)} = \{a_{1,jk}^{(k)}\}$$

    that specifies the probability of transitioning from state k to state j.
- Finally, there is the state probability matrix,

  $$\mathbf{B}^{(k)} = \{b_i^{(k)}(\mathbf{O}_{t_0,t_1})\}$$

for the set of observations where $\mathbf{O}_{t_0,t_1}$ represent the observation vector at row t_0 and column t_1. In a *continuous density* HMM, the states are characterized by continuous observation density functions. The probability density function that is typically used is a finite mixture of the form

$$b_i^{(k)}(\mathbf{O}_{t_0,t_1}) = \sum_{m=1}^{M} c_{im}^{(k)} N(\mathbf{O}_{t_0,t_1}, \mu_{im}^{(k)}, \mathbf{U}_{im}^{(k)}) \qquad (1)$$

where $N(\mathbf{O}_{t_0,t_1}, \mu_{im}^{(k)}, \mathbf{U}_{im}^{(k)})$ is a Gaussian pdf with mean vector $\mu_{im}^{(k)}$ and covariance matrix $\mathbf{U}_{im}^{(k)}$, $c_{im}^{(k)}$ is the mixture coefficient for the mth mixture in state i of super state k, $1 \leq i \leq N_1^{(k)}$.

Let $\mathbf{\Lambda}^{(k)} = \{\mathbf{\Pi}_1^{(k)}, \mathbf{A}_1^{(k)}, \mathbf{B}^{(k)}\}$ be the set of parameters that define the k^{th} super state. Using a shorthand notation, an embedded HMM is defined as the triplet

$$\lambda = (\mathbf{\Pi_0}, \mathbf{A_0}, \mathbf{\Lambda}). \qquad (2)$$

where $\mathbf{\Lambda} = \{\mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(2)}, \ldots, \mathbf{\Lambda}^{(N_0)}\}$.

This model is appropriate for face images since it exploits an important facial characteristic: frontal faces preserve the same structure of "super states" from top to bottom, and also the same left-to-right structure of "states" inside each of these "super states". The state structure of the face model and the non-zero transition probabilities of the embedded HMM are shown in Figure 1. Each state in the overall top-to-bottom HMM is assigned to a left-to-right HMM.

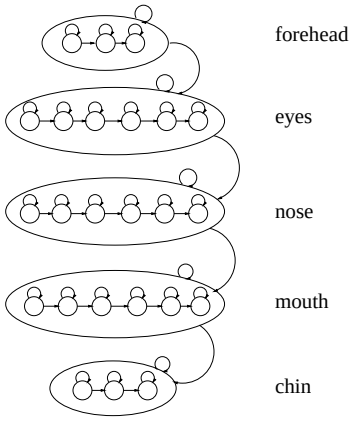

Figure 1: Embedded HMM for face recognition

The states of the embedded HMM are described by single density Gaussian pdf

$$b_i^{(k)}(\mathbf{O}_{t_0,t_1}) = N(\mathbf{O}_{t_0,t_1}, \mu_i^{(k)}, \mathbf{U}_i^{(k)}) \qquad (3)$$

and the covariance matrix is assumed to be diagonal.

3. THE OBSERVATION VECTORS

A set of overlapping blocks are extracted from the image from left to right, and top to bottom using the technique shown in Figure 2. In our embedded HMM, the observation vectors consist of six coefficients within a rectangular window (3×2) over the lowest frequencies in the 2D-DCT domain.

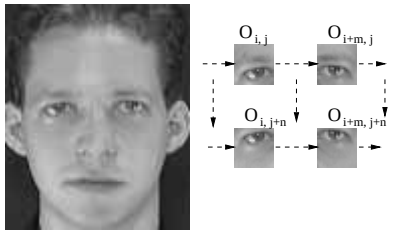

Figure 2: Face image parameterization and blocks extraction

4. TRAINING THE FACE MODELS

For face recognition each individual in the database is represented by an embedded HMM. A set of images representing different instances of the same face is used in the training set. For face detection, a set of face images representing frontal views of different individuals is used to train one face model. For both detection and recognition, the observation vectors were used to train the models as follows (Figure 3):

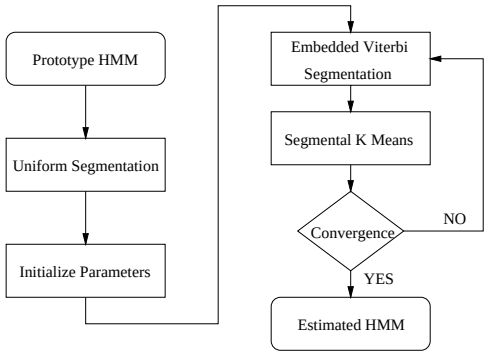

Figure 3: Training Scheme

1. First, the data is uniformly segmented to obtain initial estimates of the model parameters. The observations of the overall top-to bottom HMM are uniformly segmented in N_0 vertical super states, then, the data corresponding to each of this super states is uniformly segmented from left to right into $N_1^{(k)}$ states.

2. At the next iteration, the uniform segmentation is replaced by a doubly embedded Viterbi segmentation algorithm [6] illustrated in Figure 4.

   First, the Viterbi segmentation is applied to each row of the image, and the super state probabilities

   $$P(\mathbf{O}_{t_0,1} \ldots \mathbf{O}_{t_0,T_1}, q_{1,1}^{(t_0)} \ldots q_{1,T_1}^{(t_0)} | \lambda^{(k)}),\ 1 \leq k \leq N_0$$

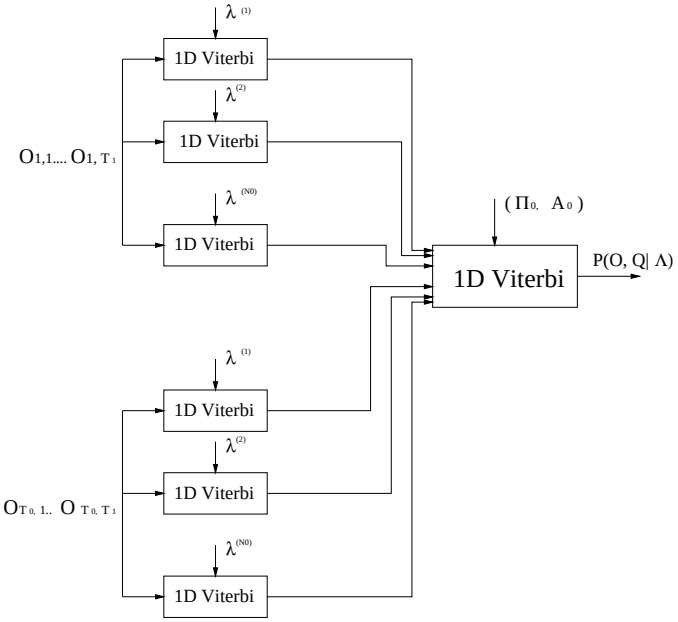

Figure 4: Doubly Embedded Viterbi Algorithm

are calculated, where $q^{(t_0)}_{1,t_1}$, $1 \leq t_1 \leq T_1$ represent the state of a super state assigned to the observation $\mathbf{O}_{t_0,t_1}$. The super state probabilities, together with the super state transition probabilities $\mathbf{A_0}$ and the initial super state probabilities $\mathbf{\Pi_0}$, are used to perform the Viterbi segmentation from the top to the bottom of the image and to determine:

$$P(\mathbf{O}_{1,1} \dots \mathbf{O}_{1,T_1}, \dots \mathbf{O}_{T_0,1} \dots \mathbf{O}_{T_0,T_1}, q_{0,1} \dots q_{0,T_0} | \lambda)$$

or using a shorthand notation $P(\mathbf{O}, \mathbf{Q}|\lambda)$. q_{0,t_0}, $1 \leq t_0 \leq T_0$ are the super states corresponding to row t_0.

3. The model parameters are estimated using an extension of the segmental k-means algorithm [6] to two dimensions.
4. The iteration stop, and the parameters of the embedded HMM are estimated, when the Viterbi segmentation likelihood at consecutive iterations is smaller than a threshold.

5. FACE RECOGNITION

The probability of the observation sequence given an embedded HMM face model is computed via a doubly embedded Viterbi recognizer. The model with the highest likelihood is selected and this model reveals the identity of the unknown face. The face recognition system has been tested on the Olivetti Research Ltd. database (400 images of 40 individuals, 10 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 of people of different ages, both males and females, showing different facial expressions, hair styles, and eye wear (glasses/no glasses). The recognition performance of the method presented in this paper is 98%. The complexity and the recognition rate of this method are compared to other HMM-based methods in Table 1.

	Recognition Rate	Complexity (additions)
HMM [2, 3, 4]	85%	$N_0^2 T_0$
HMM [8]	90-95%	$(\sum_{k=1}^{N_0} N_1^{(k)})^2 T_0 T_1$
Embedded HMM	98%	$(\sum_{k=1}^{N_0} (N_1^{(k)})^2 T_1) T_0 + {N_0}^2 T_0$

Table 1: Comparison of the recognition rate and numerical complexity (additions) for the HMM-based approaches to face recognition

Figure 5 presents some of the recognition results. The crossed images represent incorrect classifications, while the rest of images are examples of correct classification.

Figure 5: Face Recognition Results

6. FACE DETECTION

The goal of a face detection system is to locate the position of all faces in an image. A robust face detection system has to detect the faces of people, both males and females, from different races independent of their appearance (facial hair, glasses/no glasses), orientation and background. The embedded HMM structure presented in this paper allows for an efficient implementation of such a system using the doubly embedded Viterbi segmentation algorithm.

First the probability of each observation vector given a state and a super state of the model is computed. Let W_M, H_M and W_m, H_m be the numbers of observation vectors in the horizontal and vertical direction corresponding to the largest and respectively smallest face that can occur in an image, and let $\mathbf{O}_{t_0,t_1}$ be the observation vector corresponding to the left top corner of a rectangular face pattern. The doubly embedded Viterbi algorithm can be applied to each window of size $W_M \times H_M$. The total number of additions required in this case is $(W - W_m)(H - H_m)(H_M N_0^2 + H_M W_M \sum_{k=1}^{N_0} (N_1^{(k)})^2)$. To further reduce the complexity of the system, first all the super state probabilities $P(\mathbf{O}_{t_0,t_1} \dots \mathbf{O}_{t_0,t_1+W_M}, q^{(t_0)}_{1,t_1} \dots q^{(t_0)}_{1,t_1+W_M} | \lambda^{(k)})$ are

computed for all (t_0, t_1) and super states k. Second, the face likelihoods of the rectangular pattern are obtained by running the Viterbi algorithm for the overall top to bottom structure. Therefore, the total number of additions is decreased to $H(W-W_m)W_M \sum_{k=1}^{N_0}(N_1^{(k)})^2)+(W-W_m)(H-H_m)N_0^2 H_M$. No extra computation is required to determine the likelihoods of the observation vectors corresponding to the rectangular patterns included in the window of size $W_M \times H_M$ and having the left top corner at $\mathbf{O}_{t_0,t_1}$. For the template-based face detection systems, the likelihood of the patterns of different sizes has to be recomputed and consequently these methods are computationally less efficient.

The accuracy of the detection was improved by including the state duration into the overall top to bottom HMM. The duration d_i of super state i is modeled using the Poisson distribution [8] of parameter l_i. It has been shown [7] that the inclusion of the states duration increases significantly the complexity of the system. However, a very simple and efficient method is to compute the face likelihood according to:

$$log\tilde{P}(\mathbf{O}, \mathbf{Q}|\lambda) = logP(\mathbf{O}, \mathbf{Q}|\lambda) + \alpha \sum_{i=1}^{N_0} log\, p_i(d_i)$$

where α is a constant set to 1000 in our experiments. The parameter of the Poisson distribution is obtained in the training part according to:

$$l_i = \frac{number\ of\ observations\ in\ super\ state\ i}{total\ number\ of\ observations\ in\ super\ states}$$

To deal with different scales of the images in the training set and test set, the Poisson parameter is normalized to an integer value closest to:

$$\tilde{l}_i = l_i \frac{lenght\ of\ test\ sequence}{average\ length\ of\ training\ sequence}$$

The face likelihoods obtained for each rectangular pattern in the image are compared in turn to a threshold and the patterns that have likelihoods that increase this threshold are face candidates. It is natural that close patterns to have close likelihoods and therefore several patterns around the actual face location are declared to be face candidates. In order to remove these "false alarms", a face candidate represents a valid face location if its likelihood is larger than all the face likelihoods of the face candidates in a vicinity.

The face detection system proposed in this paper has been tested on 288 images of the MIT database. Each of these images has the resolution 240×256 pixels and shows one face of 16 individuals at different scales, orientations, and illuminations in a moderately cluttered background. The training set consists of 40 images at the resolution 92×112 of 40 individuals from the Olivetti Research Ltd. database. Figure 6 shows some of the detection results. The detection rate of the face detection system described in this paper is 86%.

7. CONCLUSIONS

This paper describes an embedded HMM approach for face detection and recognition that uses an efficient set of observation vectors obtained from the 2D-DCT coefficients.

Figure 6: Face Detection Results

The use of an embedded HMM model for the face is justified by the structure of the face, and also by the relatively low complexity of the model. The use of an embedded HMM increases by over 10% the recognition rate of the one-dimensional HMM and reduces significantly the computational complexity of the pseudo 2D-HMM face model introduced by Samaria [5].

In this paper a new face detection approach was introduced that uses the same face model. Preliminary results show that this approach which allows for an efficient implementation, is more flexible with respect to variations in scale, orientations and illuminations than the one dimensional HMM approach [4] or other template-based approaches.

Future work will be directed towards improving the face detection system by using the mixture density modeling of the states and by increasing the number of embedded HMMs used to model all face appearances.

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] F. Samaria and S. Young, "HMM based arhitecture for face identification," *Image and Computer Vision*, vol. 12, pp. 537–543, October 1994.

[3] A. V. Nefian and M. H. Hayes, "A Hidden Markov Model for face recognition," in *ICASSP 98*, vol. 5, pp. 2721–2724, 1998.

[4] A. V. Nefian and M. H. Hayes, "Face detection and recognition using Hidden Markov Models," in *International Conference on Image Processing*, 1998. to appear.

[5] F. Samaria, *Face Recognition Using Hidden Markov Models*. PhD thesis, University of Cambridge, 1994.

[6] 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*, vol. 16, pp. 842–848, August 1994.

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

[8] M. Russell and R. Moore, "Explicit modelling of state occupancy in Hidden Markov Models for automatic speech recognition," in *ICASSP 85*, pp. 5–8, 1985.